

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 16:34:22 ; Search time 69.24 Seconds  
(without alignments)  
1084.341 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVIINRNTYAEQEAFFRD.....LTGGLWPVVIEHVDKPSIL 434

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:
- 2: sp\_bacteria:
- 3: sp\_fungi:
- 4: sp\_human:
- 5: sp\_invertebrate:
- 6: sp\_mammal:
- 7: sp\_mmc:
- 8: sp\_organelle:
- 9: sp\_phase:
- 10: sp\_plant:
- 11: sp\_rodent:
- 12: sp\_virus:
- 13: sp\_virus:
- 14: sp\_virus:
- 15: sp\_virus:
- 16: sp\_virus:
- 17: sp\_virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2265	98.7	434	5	Q19495
2	802	35.0	418	5	Q9U318
3	716.5	31.2	461	5	Q23599
4	706.5	30.8	450	5	Q93871
5	645	28.1	458	5	Q9VMB6
6	574.5	25.0	800	5	Q21145
7	367.5	16.0	475	11	Q99P65
8	343.5	15.0	456	4	Q9UJY2
9	336.5	14.7	456	6	Q9BG85
10	328.5	14.3	456	11	Q9UJY2
11	327.5	14.3	475	4	Q9BZD2
12	320.5	14.0	428	10	Q9W5X9
13	313.5	13.7	428	10	Q64603
14	302.5	13.2	458	11	Q9DBT8
15	302.5	13.2	458	11	Q9JHF0
16	301.5	13.1	460	11	Q9JIM1

17	301.5	13.1	460	11	Q99K84
18	291.5	12.7	476	5	Q9NH21
19	291.5	12.7	476	5	Q9SRN3
20	261.5	11.4	397	4	Q9NUS9
21	249.5	10.9	408	10	Q9FWI1
22	221.5	9.7	586	5	Q9VU20
23	217.5	9.5	404	5	Q9VPP0
24	197	8.6	497	5	Q9GTP5
25	195.5	8.5	415	6	Q9BG84
26	195	8.5	384	5	Q93690
27	195	8.5	491	5	Q76343
28	187	8.2	491	5	Q76269
29	185.5	8.1	285	4	Q9BWI2
30	182.5	8.0	668	5	Q961M5
31	182	7.9	361	4	Q96PB2
32	181	7.9	301	4	Q96R00
33	181	7.9	501	5	Q9N9R1
34	181	7.9	549	5	Q95YT5
35	180	7.8	143	5	Q20396
36	170	7.4	499	5	Q9NBV4
37	161	7.0	417	10	Q944P0
38	151	6.6	463	5	Q9U763
39	147	6.4	417	10	Q9SR64
40	144	6.3	418	10	Q9M0Y3
41	142	6.2	462	5	Q9TVQ1
42	142	6.2	463	5	Q95Z04
43	141	6.1	417	10	Q944N9
44	138.5	6.0	418	10	Q9M0Y2
45	136	5.9	463	5	Q9Y010

ALIGNMENTS

RESULT 1  
Q19495 PRELIMINARY; PRT; 434 AA.  
AC Q19495  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 49.3 KDA PROTEIN.  
GN F16H11.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.;"  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX Wu X.;  
RT "The sequence of C. elegans cosmid F16H11.1;"  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX Waterston R.;  
RT "Direct Submission.;"  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U55376; AAA98003.1;  
DR InterPro: IPR002259; DER\_Nucleoside\_tran.  
DR Pfam: PF01733; Nucleoside\_tran; 1.  
DR ProDom: PD005103; DER\_Nucleoside\_tran; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 434 AA; 49317 MW; 669478D51833965 CRC64;

X not good date

Query Match 98.7%; Score 2265; DB 5; Length 434;  
Best Local Similarity 99.1%; Pred. No. 2e-177; Indels 0; Gaps 0;  
Matches 430; Conservative 0; Mismatches 4;

Qy 1 MYIINRNTYAVEQEAFFPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDG 60  
Db 1 MYIINRNTYAVEQEAFFPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDG 60

Qy 61 VETWYSKEFMGSLTIGSQLPNASINVENFLIIAGPLIYRVFAPVCFNIVNLTILIVI 120  
Db 61 VETWYSKEFMGSLTIGSQLPNASINVENFLIIAGPLIYRVFAPVCFNIVNLTILIVI 120

Qy 121 VLEPTEDSMFVFWTLCMATISINENGLYNSVGVGDPEPHTYIGALLIGNNICGLLI 180  
Db 121 VLEPTEDSMFVFWTLCMATISINENGLYNSVGVGDPEPHTYIGALLIGNNICGLLI 180

Qy 181 TVVKGIVTYFLNDEPKLVAIVYFGLSVLLVCAIALFFITKQDFYHHQKGMIREKA 240  
Db 181 TVVKGIVTYFLNDEPKLVAIVYFGLSVLLVCAIALFFITKQDFYHHQKGMIREKA 240

Qy 241 ETRDPSILMTTFTNCYQGLFNWFCFAVTLTIFFVMMVTTRGDSGFLNKMSENDEI 300  
Db 241 ETRDPSILMTTFTNCYQGLFNWFCFAVTLTIFFVMMVTTRGDSGFLNKMSENDEI 300

Qy 301 YTLTSLFVNLFAAIGSIVASKIHWPTPRYLKFAIILRALFIPFFFCNRYVQTRAYPV 360  
Db 301 YTLTSLFVNLFAAIGSIVASKIHWPTPRYLKFAIILRALFIPFFFCNRYVQTRAYPV 360

Qy 361 PFESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHSRFAAQLSVCTLMVGLLTGLW 420  
Db 361 PFESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHSRFAAQLSVCTLMVGLLTGLW 420

Qy 421 PVVIEHFVDKPSIL 434  
Db 421 AVVIEHFVDKPSIL 434

RESULT 2

Qy 903L8 PRELIMINARY; PRT; 418 AA.  
AC Q903L8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE C47A4.2 PROTEIN.  
GN C47A4.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mortimore B.J.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; 282263; CAB62793.1; -;  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 2.  
SQ SEQUENCE 418 AA; 46333 MW; 5CBDD27E06B9FC8D CRC64;

Query Match 35.0%; Score 802; DB 5; Length 418;  
Best Local Similarity 37.6%; Pred. No. 9.5e-58;  
Matches 156; Conservative 84; Mismatches 135; Indels 40; Gaps 2;

Qy 3 IINRNTYAVEQEAFFPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDGVE 62  
Db 30 IVDGADSDDEHNELLPEKGRIVFWIILLNGIGVLLPWNMFITIAPEYYVNYWFKPDGVE 89

Qy 63 TWYSKEFMGSLTIGSQLPNASINVENFLIIAGPLIYRVFAPVCFNIVNLTILIVI 122  
Db 90 THYADSFQSGVVAQVPLNVAIINVLNIRGLLYRVLAPLAFNLSLLIIVILALVIFQ 149

Qy 123 EPTEDSMFVFWTLCMATISINENGLYNSVGVGDPEPHTYIGALLIGNNICGLTI 182  
Db 150 QPSDQARMEFYIVSLIIVMAMNASGLYQNSFEGMAADFPKYSNAVIGTICGTTFSV 209

Qy 183 VKIGTYFLNDEPKLVAIVYFGLSVLLVCAIALFFITKQDFYHHQKGMIREKAET 242  
Db 210 LAIVATLAFSQAENVALIYFGLSVLLVCAIALFFITKQDFYHHQKGMIREKAET 250

Qy 243 DRPSPSILMTTFTNCYQGLFNWFCFAVTLTIFFVMMVTTRGDSGFLNKMSENDEI 302  
Db 251 -----CWLOICVFLVYFVLSVFPVLAGFQPGYTFEPN-----DVA 289

Qy 303 LTLTSLFVNLFAAIGSIVASKIHWPTPRYLKFAIILRALFIPFFFCNRYVQTRAYPV 362  
Db 290 GIAVFLNFFFAAGVAAATFVTFEPRLIIVPCVIRLILFIPFFFMFSNLYPHSRMTGVLF 349

Qy 363 ESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHSRFAAQLSVCTLMVGLLTG 417  
Db 350 TNEWIFFGNTLLAFTSGTSSGLMMYTPRCVCPPEYSKLAGQVSAVSLVIGITAG 404

RESULT 3

Qy 903L8 PRELIMINARY; PRT; 461 AA.  
AC Q23599;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE ZK809.4 PROTEIN.  
GN ZK809.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dobson R.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; 268303; CAA92642.1; -;  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 1.  
SQ SEQUENCE 461 AA; 52050 MW; CD0F115AF6A12A76 CRC64;

Query Match 31.2%; Score 716.5; DB 5; Length 461;  
Best Local Similarity 35.8%; Pred. No. 1e-50;  
Matches 163; Conservative 86; Mismatches 167; Indels 39; Gaps 9;

Qy 7 SNTYAVEQEA-FPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPD----- 59  
Db 11 NKTKKVEEESPEPEKGNLVFVILLHIGTILMPNMLTISYDFYESYKMLANLTI 70

Qy 60 -----GVETWYSKEFMGSLTIGSQLPNASINVENFLIIAGPLIYRVFAPVCFNIVNLT 114  
Db 71 TGVVTGYTYVSSNFQSTIASQVFNLLNLLNIFVIVKGLASRI--TVGLSIVAVCV 128

```
QY 115 ILILVIVLEPTEDSMNFFWVTLTGMATSNFNSGLYNSVYGVGDFFPHYTGALLIGN 174
Db 129 ITTMMFIYVETSTWLTGFTTLITIIIVLNGANGVQNSIFGLASELPFKYTNVAILGN 188
QY 175 ICGLLITVWKI--GVYFLNDEPKVAIVFGISLVILLVCATALFTTKQDYHYHQ 231
Db 189 LCGFTVLLSMTKAVTRNILD---SFAYFSIALITLVFCFISFILKKQRYQYST 244
QY 232 KGMIREK--AETDRSPSILMTTNCYQGLFNWVFCFAVTLTTP-VNMTVTTGRDS 287
Db 245 RAERQRKNDEAVDSEKGVANIATFEAPQLINLVLFVFTLISIPGVMMYVKDEKG 304
QY 288 G-----FLNKIMSENDEI----YTLTSLFNLFAAIGSIVASKIHWPTPYLK 333
Db 305 GVIDFPLPROFFINSLTHNREFSEKVFMDVTTFLQFNVPFAIGSIVAGRKQWPAPNKLW 364
QY 334 FAITLRALEFIPFEFCNRYQVTRAYPVFFESTDFVIGGIMASFGYLSALAMGYTPNV 393
Db 365 IPVYLRLLYIPFFIFCNVLPETRLPVPFFESTWLFVIAAASMSFGSYFGLAMMYTSKT 424
QY 394 VPSHYSRFAQLSVCTLMVGLLTGGLPVPVIEHFV 428
Db 425 VDPKSAQVAGMAGFELISGIVSLIETWIKMVV 459

RESULT 4
Q93871 ID Q93871 PRELIMINARY; PRT; 450 AA.
AC Q93871;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE K09A9.3 PROTEIN.
GN K09A9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z79601; CAB01882.1; -.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 450 AA; 50495 MW; CF3973D5EE517909 CRC64;

Query Match 30.8%; Score 706.5; DB 5; Length 450;
Best Local Similarity 36.0%; Pred No. 6.7e-50;
Matches 157; Conservative 88; Mismatches 164; Indels 27; Gaps 8;

QY 12 VEQEA-PPRKYNIVYWLVLGVGLPNNMFTTIAPEYVNVWF-----KPD 59
Db 21 VEEETPEPKGNLVFVILLHGIGTLMPNNMLTISVDYFESYKMLANSTIDMDTKVT 80
QY 60 GVEWYSEKFGSLTISQSLPNASINFNFLITAGLPIYVFPVPCFNIVNLITILV 119
Db 81 GDTFVYSNFSQFOTIASQVFNLLNLNIFVYKGLAGRI--TVGLSIVAVCVITMI 138
QY 120 IVLEPTEDSMNFFWVTLTGMATSNFNSGLYNSVYGVGDFFPHYTGALLIGNICGLL 179
Db 139 FIYVETSTWLTGFTTLITIIIVLNGANGVQNSIFGLASELPFKYTNVAILGNLCGTF 198
QY 180 ITVVKI---GVYFLNDEPKVAIVFGISLVILLVCATALFTTKQDYHYHQKGMIEI 236
```

```
Db 199 VTLSSMTKAVTRNILD---SFAYFSIALITLVFCFISFILKKQRYQFYSTRABEQ 254
QY 237 R---EKAETDRSPSILMTTNCYQGLFNWVFCFAVTLTTP-VNMTVTTGRDGSGLNK 292
Db 255 RAKNEEAADNEGKMANIYATEKEAPQLINLVLFVFTLISIPGVMMYVKDEKKGTYDF 314
QY 293 IMSNDEIYTLTSLFNLFAAIGSIVASKIHWPTPYLKFAITLRALEFIPFEFCNVR 352
Db 315 PLQN--YFMDVTTFLQFNVPFAIGSIVAGRKQWPAPNKLIPVYLRLLYIPFFIFCNV 372
QY 353 VQTRAYPVFFESTDFVIGGIMASFGYLSALAMGYTPNVVPSHYSRFAQLSVCTLMV 412
Db 373 PETRLPVPFFESTWLFVIAAASMSFGSYFGLSNMYSKTVDPSKAQVAGMAGFELIS 432
QY 413 GLLTGGLPVPVIEHFV 428
Db 433 GIVSGLIFTWIKFVV 448

RESULT 5
Q9VMB6 ID Q9VMB6 PRELIMINARY; PRT; 458 AA.
AC Q9VMB6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CG11045 PROTEIN (CH12067P).
GN BEST:LD04971 OR CG11045.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Eohydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hariri N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang G., Zhao Q.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
```



```
SQ SEQUENCE 475 AA; 51719 MW; F593D5D03C6CBB52 CRC64;

Query Match 16.0%; Score 367.5; DB 11; Length 475;
Best Local Similarity 25.8%; Pred. No. 3.9e-22;
Matches 123; Conservative 83; Mismatches 194; Indels 77; Gaps 15;

Qy 7 SNTAVFOEAP-----PRDKYINIVYLVLGVGVLPLPNMFIITIAPEYV 52
  |||
Db 21 SNHQEADQEAALLKLLDYPAPGLQRPEDRFGAVIIFCGLIGGLLPWNFFVT-AKEVW 79
  |||
Qy 53 NYWKPGDGVETWYSKE-----PMGSLTIGSQLPNASINFNFLIAGPLIIVRVF 102
  |||
Db 80 ---YKLRNCSSPASGEDPEDMDILNYFESYLAVASTVPSLLFLVAFNLLVNRVQVHVRVL 136
  |||
Qy 103 APVCFNIVNLTIIILVILEPTEDSMW---FEWVTILGMATISFNGLYENSIVGVGG 159
  |||
Db 137 ASLSVSLAIFVMIVLVK-----DTSWTRGFESLIACMAIISSTIFNSVYGLTG 191
  |||
Qy 160 DPHYTGALLIGNNICGLLTVVYKIGVYFLNDEPKLVAIVYFGISLVILLVCAIALFF 219
  |||
Db 192 SFPMRNAQALISGAMGG-TVSAVALLDLAASDVRDSTLAFFLMAAVFLGLCMGLYLL 250
  |||
Qy 220 ITKODFYHHQKWEIREKAETDRPS---PS-----ILWTFITNCY 258
  |||
Db 251 LSQLEYARYMRVPVPRVFSGEDNPDSAPASVAPASRVMTHTPLGPLIKKT----- 305
  |||
Qy 259 GOLFNWVFC-----FAVTLTTPVNMVTVTRGDSGLNKMSENDEIYTLTSLFVNLF 314
  |||
Db 306 ---ASLGFCAVSLYFVTAIIPALISTNIQSMHKTGSPWTS---KFFVPLVFLFNAD 359
  |||
Qy 315 AIGSIVASKIHWPPR--YLKFAILRALFIPFFFCNRYQTRAYPVFFESTDIF-VIG 371
  |||
Db 360 LCGRQVTAIWQVPGSKLLPGLVSRFCVLPVFLLCNYPQSRSLTKVLFQS-DIYPLVF 418
  |||
Qy 372 GIANSFSGYLSALAMGYTPNVPSHSRFAAQSVCTLMVGLTGLWLPVVIHFV 428
  |||
Db 419 TCLLGLSNGYLSLVLYGPKIVRELAEATSVVMLFMSVGLMGSACAALLEHFI 475

RESULT 8
Q9UJY2 PRELIMINARY; PRT; 456 AA.
AC Q9UJY2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 1 (SOLUTE CARRIER FAMILY 29
DE (NUCLEOSIDE TRANSPORTERS), MEMBER 1) (SIMILAR TO SOLUTE CARRIER FAMILY
DE 29 (NUCLEOSIDE TRANSPORTERS), MEMBER 1).
GN ENT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Graham K.A., Coe I.R., Carpenter P., Baldwin S.A., Young J.D.,
RA Cass C.E.;
RT "Genomic Sequence of the human Equilibrative Nucleoside Transporter 1
RT (hENT1).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE-COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA TISSUE-MUSCLE, RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190884; AAF02777.1; -;
```

```
DR EMBL; BC001382; AAH01382.1; -.
DR EMBL; BC008954; AAH08954.1; -.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPRT.
DR ProDom: PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 456 AA; 50219 MW; F18535A95DEBC95D CRC64;

Query Match 15.0%; Score 343.5; DB 4; Length 456;
Best Local Similarity 24.3%; Pred. No. 3.4e-20;
Matches 112; Conservative 77; Mismatches 212; Indels 59; Gaps 13;

Qy 18 PRDKYINIVYLVLGVGVLPLPNMFIITIAPEYVYVWFKPDGVE--TWYSKPEFMGSLT 74
  |||
Db 7 PQDRKAVKWLIFFMGLGLTLLPNFFMT-ATQYFTNRLDMSQNVSLVTAELSKDAQASAA 65
  |||
Qy 75 IGSOLP--NASINFNFLIAGPLIIVRFA-----PVCFNIVN-----LTIILI 117
  |||
Db 66 PAAPLPERNSLSAIFNNVMTLCAMLPLLLFTYLSFLHORIPQSVRLIGSLVAILLVELI 125
  |||
Qy 118 LVIVLEPTEDSMWFFVVTILGMATISFNGLYENSIVGVGDFFHTYIGALLIGNNICG 177
  |||
Db 126 TAILVKVOLDALP-FFVITMIKIVLINSFGAILQGSULFGLAGLPPASTAPIMSGQGLAG 184
  |||
Qy 178 LLITWVKIGVYFLNDEPKLVAIVYFGISLVILLVCAIALFFITKQDFYHHQ----- 231
  |||
Db 185 FFASVAMI-CAIASGSELSESAFGYFTACAVIILTIICYLGLPRLEFYRYQQLKEGP 243
  |||
Qy 232 -----KGMIREKAETDRPSILWTT-----FTNCYQOLFNVWFCFAVT 271
  |||
Db 244 GEQETKLDLISKGEPRAGKEESVSVNSQPTNESHISIKAILKNISVLAFSVCFITIT 303
  |||
Qy 272 LTIFPVNMVTVTRGDSGLNKMSENDEIYTLTSLFVNLFNFAAIGSIVASKIHWPP--TP 329
  |||
Db 304 IGMFPVAVTVKYSIAG-----SSTWERYFIPVSCFLTNIFFDLGRSLTAVFMWPGKDS 358
  |||
Qy 330 RYLFAILRALFIPFFFCNRYQTRAY-PVFFESTDIFVIGGIAMSFSGYLSALAMG 388
  |||
Db 359 RWLPSLVLARLVFVPLLLCN--IKPRYLTVFEHDANWAFIFMAAFSAFNGYLSALCMC 416
  |||
Qy 389 YTPNVPSHSRFAAQSVCTLMVGLTGLWLPVVIHFV 428
  |||
Db 417 FGPKKVPAAETAGATMAFFLCLGLALGALGAVSFSLFRAIV 456

RESULT 9
Q9BG85 PRELIMINARY; PRT; 456 AA.
AC Q9BG85;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 16, Last annotation update)
DE NEMPR-INSENSITIVE NUCLEOSIDE TRANSPORTER EI.
GN ENT2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu S.K., Ann D.K., Kim K., Lee V.H.L.;
RT "Molecular cloning and functional characterization of equilibrative
RT nucleoside transporters (rBENT2 and its constitutively spliced variant
RT rBENT2A) in rabbit.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323951; AAK11605.1; -.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPRT.
DR ProDom: PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 456 AA; 49858 MW; C6D55FC9D9D2A392 CRC64;
```



```
Db 80 ---FKLRSSSPATGEDPEGSD--ILNFYSYLAIVASTVPSMCLCLVANFLNVRVAVHIR 134
Qy 101 VFAPVCFNIVNLTIILIVIVLEP--TEDSMW---FFWVTLGMATSNFNSGLYENSVY 155
Db 135 VLA-----SLTYLAIFWITALKVKYDTSWTRGFPAVIVCMVILSGASTVFSSSIY 187
Qy 156 GVGDFPHYTGALLIGNNICGLITVYKIGVTVFLNDPKLVAIVYFGISLVILLVCAI 215
Db 188 GMTGSFMRNSQALISGAMGTVSASVL-VDLAASSDVRSALAFELTATIFLVLCMG 246
Qy 216 ALFETTKODEVHY-----HHQGMREIKAEKTDREPSI-----LWTTTN 256
Db 247 LYLRLSREUYAYMRPVLAHVSGBE--ELPQDSLAPSVASRFIDSHPPRPILKK 304
Qy 257 CYGOLFNFVWCFVAVTLTIFFVPMVMTTTRGDSGFLNKIMSE--NDEIYTLTSLFVFNLFA 314
Db 305 TASLGFCTVYVFFITSLIYPAVCT-----NIESLNKSGSLWTYKFFPLTFLLYNEAD 359
Qy 315 AIGSIVASKIHWPPT--RYLKFAILRALTFPFFFCNRYVOTRAYPVFFSTDFIVGG 372
Db 360 LCGRLTAWIQVPGPNRKALPGFVLLRCLTLPFLVLCNQPRVHLKTVVFGSDVYPALLS 419
Qy 373 IAMSFSCHYLSALAGYTPNVPVSHYSRFAQLSVCTLMVGLLTGGLWPVVIHFV 428
Db 420 SLGLSGNYLTSLALLYGPKIYPRELADEATGVMSFFYCLGLTLGSACTLLVHLI 475

RESULT 12
Q9MSX9
ID Q9MSX9 PRELIMINARY; PRT; 428 AA.
AC Q9MSX9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE NUCLEOSIDE TRANSPORTER.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Li J., Wang D.;
RT "Cloning and in vitro expression of the cDNA of a putative nucleoside
RT transporter from Arabidopsis thaliana.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF220759; AAF26446.1;
DR InterPro: IPR002106; AA_trna_ligase_II.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran; 1.
DR PRINTS: PR01130; DERENTRNSPT.
DR PRODOM: PD005103; DER_Nucleoside_tran; 1.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II.2; UNKNOWN_1.
SQ SEQUENCE 428 AA; 46718 MW; 1433D916FAC93CBI CRC64;

Query Match 14.0%; Score 320.5; DB 10; Length 428;
Best Local Similarity 22.9%; Pred. No. 2.4e-18;
Matches 103; Conservative 80; Mismatches 169; Indels 97; Gaps 13;

Qy 18 PRDKYNIYWLIVLGVGVLPLPWNMTIAPYVYVYVWFKPDGVETWYSKDFMGSITIGS 77
Db 35 PSDSYHFAYIIYFTLVGVGFLPLPWNMTIAPV--YFSYLYPSTAVDRIFA----- 81
Qy 78 QLPNASINVFNLFLIIA-----VYMLVAPVCLFVIVVYFAHKSASFRLNGLLFFVIALV 122
Db 82 -----VYMLVAPVCLFVIVVYFAHKSASFRLNGLLFFVIALV 122
Qy 123 EPTED-----SMSWFFWVTLGMATSNFNSGLYENSVYGVGDFPHYTGALLIGNN 174
Db 123 VPVLDLVYKQVGLYAGFDVTSRAVALSGLDALMQGGLIGVAGEMPRTQAVAGTA 182
Qy 175 ICGLLITVVKIGVTVFLNDEP---KLVAIVYFGISLVILLVCAIALFFITKQDFYHHQ 231
```

```
Db 183 GSGVLVSLRLILTRAVYQDPDGLRKRSANLYFAVGIVVWVICAIFYVAHKLPLVKKFHEE 242
Qy 232 KGME--IREKAE-----TDRSPSILWTTFFNCYQOLFNVHFCFAVTLTIIPPVMVTVTR 284
Db 243 RKNEELIREKSEKSLTGLAWRTTLDIVTKVKSHGFGIVLLYVMTLSIFFGYIT---- 298
Qy 285 GDSGFLNKIMSENDEIYT-LLTSE-----LVFNLFAAIGSIVASKIHWPPTPRYLKFAII 337
Db 299 -----EDVHSELLTDWYPILLIAAYNVFDLVGKCLTAVFMLEDEKIAVGSGI 345
Qy 338 LRALPIPFPPFCNRYVOTRAYPVFPFESTDFIVGIGIAMSFSHGYSALAMGYTPNVVESH 397
Db 346 ARLLFYPLFWGCLHG-----PMFLRTEIPVTILTCLLGLTNGYLTSLVLMILAPKSVPLR 399
Qy 398 YSRFAQLSVCTLMVGLLTGGL--WPVVI 424
Db 400 HSETAGIVTVTFLVVGSLASGSVIAFWVI 428

RESULT 13
O64603
ID O64603 PRELIMINARY; PRT; 428 AA.
AC O64603;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F1707.13 PROTEIN.
GN F1707.13.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Oji O.,
RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,
RA Conway A.R., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,
RA Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,
RA Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F1707 sequence.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC003671; AAC18807.1;
DR InterPro: IPR002106; AA_trna_ligase_II.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran; 1.
DR PRODOM: PD005103; DER_Nucleoside_tran; 1.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II.2; UNKNOWN_1.
SQ SEQUENCE 428 AA; 46764 MW; B762E36C07516DB9 CRC64;

Query Match 13.7%; Score 313.5; DB 10; Length 428;
Best Local Similarity 23.2%; Pred. No. 9.1e-18;
Matches 100; Conservative 84; Mismatches 186; Indels 61; Gaps 11;

Qy 18 PRDKYNIYWLIVLGVGVLPLPWNMTIAPYVYVYVWFKPDGVETWYSKDFMGSITIGS 77
Db 35 PSDSYHFAYIIYFTLVGVGFLPLPWNMTIAPV--YFSYLYPSTAVDRIFAIVYM----- 85
Qy 78 QLPNASINVFNLFLIIA-----GLIYRVFAPVCFNIVNLTIILIVIVLEPTEDSMSWF 132
Db 86 -----LVALVCLFVIVVYFAHKSASFRLNGLLFFVALLVVPVLDLVYKQVGLYAG 140
Qy 133 FWTIGMATSNFNSGLYENSVYGVGDFPHYTGALLIGNNICGLITVVKIGVTVFLN 192
Db 141 FDTVSAVALSGLDALMQGGLIGVAGEMPRTQAVAGTAGSGVLVSLRLILTRAVY 200
```





Search completed: July 23, 2002, 16:42:38  
Job time: 496 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 16:35:07 ; Search time 24.68 seconds  
(without alignments)  
680.887 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVIIINRNTYAVEQEAFPRD.....LTGGLWPVVEHVDKPSIL 434

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	343.5	15.0	455	1 ENT1_HUMAN	Q9808 homo sapien
2	326.5	14.2	456	1 ENT2_HUMAN	Q14542 h equilibra
3	320.5	14.0	456	1 ENT2_RAT	O54699 rattus norv
4	301	13.1	456	1 ENT1_RAT	O54698 rattus norv
5	209.5	9.1	327	1 ENT2_MOUSE	Q61672 m equilibra
6	182.5	8.0	17	1 FU26_YEAST	P31381 saccharomyc
7	134	5.8	432	1 V672_METJA	Q58086 methanococc
8	124.5	5.4	506	1 HITB_HAEIN	P71338 haemophilus
9	117.5	5.1	559	1 YAE8_SCHPO	Q9852 schizosacch
10	113.5	4.9	477	1 Y983_MOUSE	P47344 mycoplasma
11	113.5	4.9	587	1 T9S3_MOUSE	Q9et30 mus muscula
12	113.5	4.9	589	1 T9S3_HUMAN	Q9hd45 homo sapien
13	113.5	4.9	604	1 NUSM_HORSE	P48656 equus cabal
14	113	4.9	459	1 TCR2_BACSU	P14512 bacillus su
15	113	4.9	614	1 YDNK_LACIC	P42377 lactococcus
16	112	4.9	383	1 NAPA_ENTHR	P26235 enterococcu
17	111.5	4.9	397	1 TCR2_HUMAN	P55085 homo sapien
18	111.5	4.9	433	1 TCR2_MOUSE	P02983 staphylococ
19	111.5	4.9	439	1 UHPC_ECOLI	P09836 escherichia
20	110.5	4.8	479	1 Y988_MYCPN	P75535 mycoplasma
21	110.5	4.8	785	1 ISP4_SCHPO	P40900 schizosacch
22	108.5	4.7	442	1 UHPC_SALTY	P27669 salmonella
23	108	4.7	388	1 YUBA_BACSU	Q32086 bacillus su
24	106.5	4.6	399	1 PAR2_MOUSE	P55086 mus muscula
25	106	4.6	547	1 NUSM_ASCSU	P24884 ascaris suu
26	105.5	4.6	598	1 THIX_YEAST	Q08485 saccharomyc
27	105	4.6	542	1 ATN1_YEAST	P13090 saccharomyc
28	104.5	4.6	894	1 YN86_YEAST	P27514 saccharomyc
29	104	4.5	527	1 NU2M_ACACA	Q37376 acanthameob
30	104	4.5	552	1 NU5M_RHISA	Q9zym7 rhinoceph
31	103.5	4.5	345	1 NUOH_RHOCA	P42032 rhodobacter
32	103.5	4.5	606	1 NU5M_EQASU	P92485 equus asinu
33	103	4.5	459	1 NU4M_BOVIN	P03910 bos taurus

RESULT	ENT1_HUMAN	1
ID	ENT1_HUMAN	STANDARD; PRT; 455 AA.
AC	Q9808;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Equilibrative nucleoside transporter 1 (Equilibrative	
DE	nitrobenzylmercaptopurine riboside-sensitive nucleoside transporter)	
DE	(Equilibrative NMPR-sensitive nucleoside transporter) (Nucleoside	
DE	transporter, es-type).	
GN	SLC29A1 OR ENT1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.	
RC	TISSUE=Placenta;	
RX	MEDLINE=97140266; PubMed=8986748;	
RA	Griffiths M., Beaumont N., Yao S.Y.M., Sundaram M., Boumah C.E.,	
RA	Davies A., Kwong F.Y.P., Coe I., Cass C.E., Young J.D., Baldwin S.A.;	
RT	"Cloning of a human nucleoside transporter implicated in the cellular	
RL	uptake of adenosine and chemotherapeutic drugs."	
RP	Nat. Med. 3:89-93(1997).	
RP	[2]	
RC	SEQUENCE FROM N.A.	
RA	Lum P.Y., Ngo L.Y., Bakken A.H., Unadkat J.D.;	
RT	"Critical structural determinants for high affinity binding of	
RL	nucleosides to the equilibrative NMPR-sensitive nucleoside	
RT	transporter (es) cloned from the human jejunum."	
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.	
CC	-1- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS	
CC	THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO	
CC	LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYL MERCAPTOPYRINE	
CC	RIBOSE (NMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER	
CC	AFFINITY FOR ADENOSINE. INHIBITED BY DIPYRIDAMOLE AND DILAZEP	
CC	(ANTICANCER CHEMOTHERAPEUTICS DRUGS).	
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.	
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, MAMMARY GLAND,	
CC	ERYTHROCYTES AND PLACENTA, AND ALSO IN FETAL LIVER AND SPLEEN.	
CC	-1- PTM: GLYCOSYLATED.	
CC	-1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; U81375; AAC51103.1; "	
DR	EMBL; AF079117; AAC62495.1; "	
DR	MI; 602193; "	



```
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 124 144 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 130 MISSING (IN ISOFORM HNP36).
FT CONFLICT 200 200 Y -> C (IN REF. 1).
SQ SEQUENCE 456 AA; 50172 MW; AEF1B244397508E1 CRC64;

Query Match 14.2%; Score 326.5; DB 1; Length 456;
Best Local Similarity 22.9%; Pred. No. 1.7e-15;
Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

Qy 18 PRDKNYIVWLVLVGVGVLPPNWNFTIAPEYV-----VNYWEKPDGVE 62
Db 7 PRDSYHLVGVISFFILGLGTLPLPWNFFTAIPYQARLAGAGNSTARILSTNHTGPDFAF- 65
Qy 63 TWYSKEFGSLTIGSQPNASINVFNLFIITAGPLIYRVFAPVCNFIIVNLTIIILVIL 122
Db 66 -----NENNVWTLSSQLPPLLFTLLNSFLYQCVETVRLG-----SLLAIIILLFALTAAL 116
Qy 123 EPTEDSMWFVWTLGMATSNFNSGLYENSYGVGGDPHTYTGALLIGNICGLLTV 182
Db 117 VKVDMSPGPFFSTMASVCYNSFNSAVLQSLFGLQGLTPTSTYTLFSLGGLAGIFAAL 176
Qy 183 VKI-----GVTYFLNDEPKLVAIYF---GISLVILLVCATALEFIKQDPY---HYHHQ 231
Db 177 AMLLSMASGV-----DAETSALGYFITPYVGLMSIVCYLSLPHLFARYILANKSSQA 230
Qy 232 KGBIREKA-----ETDRPSPSILWTTFNC 257
Db 231 QAOELTKAELQSDENGIPSPQKVALTLDDLLEKPESEPDPEQKPGKPSVTFVQKI 290
Qy 258 YGQLFNWFCAVTLTTPVVMVTTRGDSGLNKMSENDEIYTLTSLTFVFNLFRAIG 317
Db 291 WLTAICLVLTFTVLSVFPATATVTS-----PKWQSQFFNPICCFLLNFMWLWG 344
Qy 318 SIVASKTHWP--TPRYLKFAILRALTFPFFFCNYRVOTRAYPVFEESTDIFVIGGIA 375
Db 345 RSLTSYFLWPDDESRLLPLLVCLRFVPLFVPLMCHVQPSR-LPILFPQDAYFITFMLLF 403
Qy 376 SFSHGYSALAMGYTPNVVPSHSRFAQLSVCTLMVGLLTG 417
Db 404 AVSNGYLVSLTMCLAPQVLPHEREVAGALMTFFFLALGLSCG 445

RESULT 3
ENT2_RAT
ID ENT2_RAT STANDARD; PRT; 456 AA.
AC OS4699;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Equilibrative nucleoside transporter 2 (Equilibrative
DE nitrobenzylmercaptopyrine riboside-insensitive nucleoside transporter)
DE (Equilibrative NBMPR-insensitive nucleoside transporter) (Nucleoside
DE transporter, ei-type).
GN SLC29A2 OR ENT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Jejunum;
```

```
RX MEDLINE=98019212; PubMed=9353301;
RA Yao S.Y.M., Ng A.M.L., Muzyka W.R., Griffiths M., Cass C.E.,
RA Baldwin S.A., Young J.D.;
RT "Molecular cloning and functional characterization of
RT nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive
RT (el) equilibrative nucleoside transporter proteins (rENT1 and rENT2)
RT from rat tissues.";
RL J. Biol. Chem. 272:28423-28430(1997).
CC -1- FUNCTION: MEDIATES BOTH INFUX AND EFFLUX OF NUCLEOSIDES ACROSS
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS INSENSITIVE (EI)
CC TO LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMERCAPTOPURINE
CC RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. SPECIFIC FOR
CC NUCLEOSIDES, BUT MAY ALSO TRANSPORT HYPOXANTHINE. MAY ALSO PLAY A
CC ROLE IN THE EFFLUX OF INOSINE AND HYPOXANTHINE FROM MUSCLE CELLS
CC DURING THE NET DEGRADATION OF PURINE NUCLEOTIDES THAT OCCURS
CC DURING STRENUOUS EXERCISE AND/OR IN THE REUPTAKE OF THESE PURINES
CC DURING THE RECOVERY PROCESS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: RESISTANT TO DIPYRIDAMOLE AND DILAZEP INHIBITION
CC (ANTICANCER CHEMOTHERAPEUTICS DRUGS).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF015305; AAB88050.1; .
CC InterPro: IPR002259; DER_Nucleoside_tran.
CC Pfam: PF01733; Nucleoside_tran; 1.
CC PRINTS: PR01130; DERENTRNSPRT.
CC ProDom: PD005103; DER_Nucleoside_tran; 1.
CC Transmembrane; Transport; Glycoprotein.
CC DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 13 29 POTENTIAL.
CC DOMAIN 30 69 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 69 93 POTENTIAL.
CC DOMAIN 94 97 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 98 116 POTENTIAL.
CC DOMAIN 117 124 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 125 143 POTENTIAL.
CC DOMAIN 144 160 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 161 185 POTENTIAL.
CC DOMAIN 186 192 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 193 213 POTENTIAL.
CC DOMAIN 214 291 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 292 311 POTENTIAL.
CC DOMAIN 312 323 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 324 342 POTENTIAL.
CC DOMAIN 343 359 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 360 378 POTENTIAL.
CC DOMAIN 379 393 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 394 413 POTENTIAL.
CC DOMAIN 414 431 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 432 452 POTENTIAL.
CC DOMAIN 453 456 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 456 AA; 50265 MW; DA97C2C578E1EE9D CRC64;

Query Match 14.0%; Score 320.5; DB 1; Length 456;
Best Local Similarity 23.5%; Pred. No. 4.4e-15;
Matches 107; Conservative 76; Mismatches 200; Indels 73; Gaps 13;

Qy 18 PRDKNYIVWLVLVGVGVLPPNWNFTIAPEYVYVNYWFKPDGVTWYSK-----EF 69
Db 7 PRDSYHLVGVISFFILGLGTLPLPWNFFTAIPYQARLAGAGNSTARILSTNHTGPDFAF 66
Qy 70 MGSLTIGSQLPNASINVFNLFIITAGPLIYRVFAPVCNFIIVNLTIIILVIVLEPTDSM 129
```

```
Db 67 NNVVTLTLLSPLFTLNSFLYQIPESVRILG-----SLLAILLFALTAALVKVDLSP 122
Qy 130 SFFFWTLTGMATSNFNSGLVNSVYGGDPHTYIGALLIGNNICGL-----LITVVK 184
Db 133 GLFFSITWASVWFNSFCNAVGLGSLFGQGLTWPSTYSILFSGGGLAGFAALAMLTSLA 182
Qy 185 IGVTYFLNDEPKLVAIVF---GISLVLLVCAIAL-----FFITKQ----- 223
Db 183 SGV-----DQTSALGVFITPCVGILLISICLYSLPHLKFAFYLYTKKPPQAVQLETK 236
Qy 224 -DFYHYHQQKMEIR-----EK-----AETDRSPSILWTFNFCYQQLNV 264
Db 237 AELLGADEKNGIPVSPQAGPTLDLDPEKELEGLGLEPQKPKPSVVFVFRKILWLTACL 296
Qy 265 WFCRAVTLTIP-VMVTVTTRGDSGLNKMSENDEITLLTSRLVENLFAAIGSIVASK 323
Db 297 VLVFVTLVSVFAITAMVTTSSN-----PKWSQFFNPICCFLLFNWMDLGRSLTSY 350
Qy 324 IHWP--TPRYLKFALILRALFIPFFFCNRYVQTRAYVFFESTDIFVIGGTAMSFSGY 381
Db 351 FLWPEDESQLLPLVLCFLFVPLFMLCHVQPAR-LPIFWQDAYFITEMLLFAISNGY 409
Qy 382 LSALAMGTPNVVPSHYSRFAQLSVCTLMVGLLTG 417
Db 410 FVSLTMCPLAQVLPHEREVAGALMTFFELALGLSCG 445

RESULT 4
ENTL_RAT STANDARD; PRT; 456 AA.
ID ENTL_RAT AC
OS 054698;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Jejunum;
RX MEDLINE=98019212; PubMed=9353301;
RA Yao S.Y.M., Ng A.M.B., Muzika W.R., Griffiths M., Cass C.E.,
RA Baldwin S.A., Young J.D.;
RT "Molecular cloning and functional characterization of
RT nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive
RT (ei) equilibrative nucleoside transporter proteins (rENT1 and rENT2)
RT from rat tissues.";
RL J. Biol. Chem. 272:28423-28430(1997).
CC -|- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO
CC LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYL MERCAPTOPYRINE
CC RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER
CC AFFINITY FOR ADENOSINE AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER
CC INHIBITION (ANTICANCER CHEMOTHERAPEUTICS DRUGS).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein.
CC -|- TISSUE SPECIFICITY: EXPRESSED IN JEJENUM, LIVER AND LUNG.
CC -|- PTM: GLYCOSYLATED.
CC -|- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AF015304; AAB88049.1; -
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPRT.
DR ProDom; PD003103; DER_Nucleoside_tran; 1.
KW Transmembrane; Transport; Glycoprotein.
FT INIT_MET 0 0
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 12 28 POTENTIAL.
FT DOMAIN 29 81 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 82 106 POTENTIAL.
FT DOMAIN 107 110 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 111 129 POTENTIAL.
FT DOMAIN 130 137 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 138 156 POTENTIAL.
FT DOMAIN 157 173 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 174 198 POTENTIAL.
FT DOMAIN 199 205 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 206 226 POTENTIAL.
FT DOMAIN 227 290 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 291 310 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 311 322 POTENTIAL.
FT TRANSEM 323 342 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 343 359 POTENTIAL.
FT TRANSEM 360 378 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 379 393 POTENTIAL.
FT TRANSEM 394 413 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 414 431 POTENTIAL.
FT TRANSEM 432 452 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 453 456 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 456 AA; 49885 MW; A34CE92C20836D9B CRC64;

Query Match 13.1%; Score 301; DB 1; Length 456;
Best Local Similarity 22.3%; Pred. No. 9.3e-14;
Matches 107; Conservative 84; Mismatches 191; Indels 98; Gaps 17;

Qy 18 PRDKNYIVWLVLVGVGVLLPWNFFITTAPEYVYVWFKPDGV-----ETWYSKEFGMS 72
Db 6 PDQRYKAVMLFFLGLGLTLLPWNFFIT-ATQYETSLRNTSQNISLVTNQSCSEALAD 64
Qy 73 LTIGQLPNASINVENFLIITAGPLIYRVEAPVCFN-----IVNLITII 115
Db 65 PSVSLPARSSLSAIFNNVMTLCAMLPLLIF--TCLNSFLHOKVQSLSRILGLLAILLVF 122
Qy 116 LILVIVLEPTEDSMSEWFFVTGLMATSNFNSGLVNSVYGGDPHTYIGALLIGNNI 175
Db 123 LVTATLVKQMDALS-FFIITMIKIVLINSFGAILQASLFGLAGVLPANYTAPIMSGOGL 181
Qy 176 CGLIITVVKI-----GWTFYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQ 223
Db 182 AGFTTSVAMICAVASGSKLSSEAFGYFIT---ACAVV-----ILAILCYLALPWM--- 228
Qy 224 DFYHYHQQ-----KMEIR-EKATDRSPSIL----- 250
Db 229 EFRHYQLNLNAGPAEQTKLDLISEGEEPRGREGSGVPGPNSLPANRNSIKAILKSI 288
Qy 251 WTTFTNCGQLFNWVFCFAVTLTIFFPVMVTVTTRGDSGLNKMSENDEIYTLTSLFVLF 310
Db 289 WVL-----ALSVCFTFTVIGLFP---AVTAEVESIACTSPWKNK-YEIPVACFLNF 337
Qy 311 NLFRAIGSIVASKIHWP--TPRYLKFALILRALFIPFFFCNRYVQTRAYVFFESTDIF 368
Db 338 NVFDWLGSRITAICMWFQGDSDRWLPVLVACRVVFIPLMLCNVK-QHHYLSFLSKHDWVF 396
Qy 369 VIGGTAMSFSGYLSALAMGVTPTNVVPSHYSRFAQLSVCTLMVGLLTGGVLPVVIEHFV 428
Db 397 ITFMAAFASNGYLASLWCMCFGPKKVPAPAEATAGNIMSFELCLGLALGALVSLFLRALV 456
```



## RESULT



FT	VARIANT	104	104	G > C (IN STRAIN TN106).
FT	VARIANT	116	116	V -> G (IN STRAIN TN106).
FT	VARIANT	167	167	S -> F (IN STRAIN TN106).
FT	VARIANT	252	252	T -> M (IN STRAIN TN106).
SQ	SEQUENCE	506 AA;	56546 MW;	ADA28861C1481A1D CRC64;

  

Query Match		5.4%;	Score 124.5;	DB 1;	Length 506;
Best Local Similarity		21.6%;	Pred. No. 0.11;		
Matches 105;		Conservative 69;	Mismatches 144;	Indels 167;	Gaps

  

QY	26	YW----	LVILVCGVLLPWNNFI	IAIEY----	YVNWFXPDGVETWYSKE-FNGSLTIGS	77
		:	: : :	:	:	:
DB	7	FWTLLTILIGLCLPFL	VXVILURATEVGL	TRSVELLFRPRMAELLSNTMLMLVCVTICA	66	
		:	:	:	:	:
QY	78	QLPNASINVENFL	LIATAGPLIVVFAPCFN	IVNLNLIILIVLIVLEPTEDSMW----	FFW 134	
		:	:	:	:	:
DB	67	----	ISLGTCAELLER----	YRFKGAFFEVAMTILCIPAFVSQFTWLSTFRVEVFW	118	
		:	:	:	:	:
QY	135	VTIGMATSI	NF-----	-----SNGLYENSYGVGGDFPHYIGAL-----	LIGNN 174	
		:	:	:	:	:
DB	119	GTIGIMTSL	FFPLAYLPVSAILKRLDRS	LEEVSL-SLGRKSPVYTFWVAISPOLKPAIGSS	177	
		:	:	:	:	:
QY	175	ICGLLIT--	VWIKGYTYFLNDEPKL	VAI-----VYFG-----	ISLVILLVCAI-----A 216	
		:	:	:	:	:
DB	178	I--LLIALHMLV	FGVAGSIUNYQTT	TAIFQYEMSFNNSTAALLSAVLMAICILIVFGE	235	
		:	:	:	:	:
QY	217	LEFRTKDEYHYHOK	MEIRAEKTRDPS	PSILWTTFNCYQ-----LFWNWFCEAV	270	
		:	:	:	:	:
DB	236	IFFRGKQTL--	YHSGKGV-----	TRP-----YLVKTLFSGKQCLTGFSSIFILSI	280	
		:	:	:	:	:
QY	271	TLAIFPYMM--	-----TVTTRGD-----	SGFLNK-IMSENDEIYTLTISF-----	307	
		:	:	:	:	:
DB	281	GV--PVIMLIY	WLIVGTSL	ESAGDFSEFLSFAFSNFIISGLGALLTVMCALPLVWAAVR	337	
		:	:	:	:	:
QY	308	-----LVENL	FPAATG-----	SIVASKTHWTPPYLKFAIILRALFI-----	343	
		:	:	:	:	:
DB	338	YRSYLTIDRLP	LYLLHAVPLGVL	IALSLVYFESHYANDLYQTFFVIIAYFMYLPLPMAQT	357	
		:	:	:	:	:
QY	344	-----	-----PEFFECNRYQ	TRAYPVFFESTDIFVIGGITAMSFSGH	380	
		:	:	:	:	:
DB	398	TLRASLEQLSD	QIEKVQSGSLGRNP	YIF-----RTLPLA-----	ILPGVAAAFALV 444	
		:	:	:	:	:
QY	381	YLSAL	385			
		:	:			
DB	445	FLNLM	449			
		:	:			

  

RESULT	9	
YAEAC_SCHPO		
ID	YAEAC_SCHPO	STANDARD;
AC	Q09852;	PRT; 559 AA.
DT	01-FEB-1996	(Rel. 33, Created)
DT	01-FEB-1996	(Rel. 33, Last sequence update)
DE	01-MAR-2002	(Rel. 41, Last annotation update)
DE	Putative inorganic phosphate transporter	C23D3.12.
GN	SPAC23D3.12	
OS	Schizosaccharomyces pombe (Fission yeast).	
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;	
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;	
OC	Schizosaccharomycetes.	
OX	NCBI_TaxID=4896;	
RP	[1]	
RN	SEQUENCE FROM N.A.	
RC	STRAN-972;	
RA	Niblett D., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;	
RL	Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.	
CC	-1- FUNCTION: HIGH-AFFINITY TRANSPORTER FOR EXTERNAL INORGANIC	
CC	PHOSPHATE (BY SIMILARITY).	
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).	
CC	-1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO	
CC	YEAST PHO84.	
CC	-----	

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: Z64354; CAA91247.1; -  
InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Hypothetical protein; Phosphate transport; Transmembrane.  
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 46 66 1 (POTENTIAL).  
FT DOMAIN 67 94 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 95 115 2 (POTENTIAL).  
FT DOMAIN 116 118 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 119 139 3 (POTENTIAL).  
FT DOMAIN 140 144 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 145 165 4 (POTENTIAL).  
FT DOMAIN 166 188 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 189 209 5 (POTENTIAL).  
FT DOMAIN 210 224 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 225 245 6 (POTENTIAL).  
FT DOMAIN 246 335 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 336 356 7 (POTENTIAL).  
FT DOMAIN 357 381 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 382 402 8 (POTENTIAL).  
FT DOMAIN 403 409 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 410 430 9 (POTENTIAL).  
FT DOMAIN 431 438 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 439 459 10 (POTENTIAL).  
FT DOMAIN 460 474 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 475 495 11 (POTENTIAL).  
FT DOMAIN 496 497 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 498 518 12 (POTENTIAL).  
FT DOMAIN 519 559 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 559 AA; 62316 MW; 2E8958F86C2092E2 CRC64;

Query Match 5.18; Score 117.5; DB 1; Length 559;  
Best Local Similarity 20.18; Pred. No. 0.35;  
Matches 101; Conservative 73; Mismatches 175; Indels 153; Gaps 25;

QY 26 YWLV-----LVFG-----VLLPNNFI--TIAPEYVNVWFKPDGVETWYSKEFMGSL 73  
DB 34 HWLGLTKREFKLMGFAGFFLDYDLFIINLVSPIVEYLYWGLEGGKKPKHPYSGIHG-- 91  
QY 74 TIGSQLPNASINVENFLIITAGPLIYR-----VFAPVCFNIVNLTILILVIVLE 123  
DB 92 -----LYNAANIGNVF-----GQILFGMGDFGKFKFYGKEMIVVIATVILVIALPKSI 142  
QY 124 PTE-DSMSWFF--WVTGLMATSNFNSGLYENSVYGVGSDPFTYI-----GA 168  
DB 143 PTLPLCKMWFARWL-LGL-----GIGGDYPMSTATISERSLLSRGT 185  
QY 169 LLI-----GNICGLLITVWKIGVYFLNDEP-----KLVAI--VYFGSLVILL 211  
DB 186 LLSIVSFQCGFTLAGAIVTILLACF-----EKPLNORGEYTKLEGVWRQMGLALVPAL 241  
QY 212 VCAI-----ALFFTTKDFY-----HYHHOKGMIREKAET-- 242  
DB 242 LVLLPLRTMKESYEQSKALNKYTDNDTYIADDEPKKDNQNVVEKQINLTSSGHP 301  
QY 243 -----DRSPSILWTTFTNCYQQLFNWPCF-----AVTLTFIPVMMFTVTRGDSGF 289  
DB 302 TSTEDFGDKRASTVPTSGTSGFIEYSQWHFKHLLATAVSWFLDIAFYGVNLQSVI 361  
QY 290 LNKI--HSENDIEVTLTSLVFNLFPAIGSIVASKTHWPT-----PRYLFAIILRALFI 343  
DB 362 LKAIFGSSGKNEYHTLMRG-AIGNLNTAIAGYVPG--YWFVTVFLVEKLGKRWIQLQGLFI 418

QY 344 PFFEFNRYVQTRAYPVFFESTDIEFVIGG-----IAMSFSHGYSALAMGYTPNVVPSH 397  
DB 419 TGLMFA-----IAGSWDTISTGGRFACFVIAOFFSNGFNATFLYPAEVFFAR 468  
QY 398 YSRFAAQLSVCTLMVGLLTGGL 419  
DB 469 VRGTAHGLSALGKCGAILASL 490  
RESULT 10  
Y098\_MYCGE  
ID Y098\_MYCGE STANDARD; PRT; 477 AA.  
AC P47344; Q49231; Q49509;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
GN Hypothetical protein MG098.  
OS MG098.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.D.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RL "The minimal gene complement of Mycoplasma genitalium.";  
RT Science 270:397-403(1995).  
RN [2]  
RP SEQUENCE OF 46'-155 AND 278-382 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bost K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
sequencing.";  
RL J. Bacteriol. 175:7918-7930(1993).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GATC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U39690; AAC71316.1; -  
EMBL: U01701; AAB01013.1; -  
EMBL: U01782; AAD12771.1; -  
TIGR: MG098; -  
DR InterPro: IPR003837; Glu-trnAGln.  
DR Pfam: PF02686; Glu-trnAGln; 1.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 23 43 POTENTIAL.  
FT TRANSMEM 61 81 POTENTIAL.  
FT TRANSMEM 108 128 POTENTIAL.  
FT TRANSMEM 131 151 POTENTIAL.  
FT TRANSMEM 171 191 POTENTIAL.  
FT TRANSMEM 228 248 POTENTIAL.  
FT TRANSMEM 285 305 POTENTIAL.  
FT TRANSMEM 326 346 POTENTIAL.  
FT CONFLICT 87 87 F -> V (IN REF. 2).  
FT CONFLICT 372 382 DMOSLIMPVNI -> EYAKLNIAOCY (IN REF. 2).  
SQ SEQUENCE 477 AA; 54045 MW; 3FCCDD95A2C35684 CRC64;

```
Query Match      4.9%; Score 113.5; DB 1; Length 477;
Best Local Similarity 19.7%; Pred. No. 0.56;
Matches 89; Conservative 70; Mismatches 130; Indels 163; Gaps 20;

QY 22 YNIVWLVILVGVGLLPNNFITTAPYYVYVWFKPGVETWYSKER-----MSSL 73
DB 20 YFAVVFLLIIIV-----YLFKKILFLFFFKRYPKNTPKIGVSGNI 58
QY 74 TIGSOLPNASINVFNLFLIIAGLIYRVFAPVCFNIVNLTIILIVILEPTEDSMWFF 133
DB 59 TTIIAMIAVAVSVVLVLMALAGGLTAALFPGYGRVLELLLVKISGL-----LF 109
QY 134 WVTLMG--ATSFNSNGLYSVGVGGDPHTYIGALLIGNNICGLL-----179
DB 110 GPLIGISFAATIDELTVIFSGGVFNIG-----YVLGAILTG-MIAGILREVLISTSFLNN 163
QY 180 -----ITVVKIG-----VTYFL-----NDE-----PK 196
DB 164 KTLDFAYLVLSVGMVFASFVLTQTFVIVSTQNLASFQSDQIVLRFNASPLNFSISLQR 223
QY 197 LVAIYVFGISVILLVCAIALFFITKQDFYHYHOKGMEIREKAETDRPSPSILWTTFN 256
DB 224 YVQIIFYFAMVVIITMVVLYFWIILKQHNAYSKFFERYKXKHANHOFTLVFL--TKEN 281
QY 257 CYGQLENVWFCAVTLTIFFVMMVTTRGDSGLNKMSENDEIYTLTSLFVFNLFPAI 316
DB 282 WFLYLNLV-----ITLAT-----TSLMINI--AF 304
QY 317 GSIVASKIHWTPRYLKFAILRALFIFFPFNCVTRVOTRAYPVFFESTDIFVIG-----371
DB 305 IPIEDTQ---TTGQFYDFWLLVRLFLAPLFLLDIIIV---IYPILLLTPIMLKGFRTVA 358
QY 372 -----GIAMSFSGVLSALAMGYTPNVVPSH 397
DB 359 SETQTKGKKFSFD--MOSLIM--PNVI-SH 384

RESULT 11
T9S3_MOUSE      STANDARD;      PRT;      587 AA.
AC Q9ET30;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transmembrane 9 superfamily protein member 3 precursor.
GN TM9SF3
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Warner S.J., Lomax M.I.;
RT "Evolution of the TM9 super family of membrane spanning proteins.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 79-587 FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -|- SIMILARITY: BELONGS TO THE NONASPANTIN (TM9SF) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF269151; AAF98160.1; -
DR EMBL; BC004799; AAH04799.1; ALT_INIT.
```

```
DR InterPro; IPR004240; EMP70.
KW Pfam; PF02990; EMP70; 1.
FT SIGNAL; Transmembrane.
FT SIGNAL 1 26
FT CHAIN 27 587
FT TRANSMEM 222 242
FT TRANSMEM 292 312
FT TRANSMEM 326 346
FT TRANSMEM 358 378
FT TRANSMEM 387 407
FT TRANSMEM 447 467
FT TRANSMEM 480 500
FT TRANSMEM 517 537
FT TRANSMEM 549 569
FT CARBOHYD 172 172
FT CARBOHYD 417 417
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 587 AA; 67544 MW; 22FD4F8588FEC2AE CRC64;

Query Match      4.9%; Score 113.5; DB 1; Length 587;
Best Local Similarity 18.2%; Pred. No. 0.69;
Matches 62; Conservative 49; Mismatches 127; Indels 103; Gaps 14;

QY 96 PLIYRVFAPVCNIVNLTIILIVILEPTEDSMWFFWVTILGMATSFNSNGLYNSVY 155
DB 286 PLIFSLLIGSGQIFAVSLIVIVAMIEDLYTERGSMSTAIFFVAATSPVNGYFGSLY 345
QY 156 GVGGD---PHTYIGALLIGNNICGLLITVVKYTYTYFLNDEPKLVAIYV-----FGI 205
DB 346 ARQGGRRWKQMFAGFLIPAMVC-----GTAFFIN---FTAIYHASRAIPFG- 391
QY 206 SLVILLVCAIALFFITKQDFY-----HYHOKGMEIREKAETDRPSPSILW---TTFIN 256
DB 392 --TWAVACCICFFVILPLNLVLTILGRNLSGQPNPCRVNA-VPRPIPEKKFMFEPVIV 448
QY 257 CYGQL-----FNWFCFAVTLTIFFVMMVTTRGDSGLNKMSEN 297
DB 449 CLGGILPFGSIFEMVFIFFTSWAYKIYVYVGMFLVLVILCIVTV-----494
QY 298 DEIYTLTSLFVFNLFPAALGSIIVASKIHWTPRYLKFAILRALFIFFPFNCVTRVOTRA 357
DB 495 --CVTIVCTYFLN-----AEDYRWQTSFLSAA--STAIYVYMYSFYFFFKTKM 541
QY 358 YPVFFESTDIFVIGGIAMSFSGVLSAL-----AMGY 389
DB 542 YGLF-----QTSFYFGYMAVFSTALGIMCGAIGY 570

RESULT 12
T9S3_HUMAN      STANDARD;      PRT;      589 AA.
AC Q9HD45; Q9UHH8; Q9NWL8; Q9P0G9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transmembrane 9 superfamily protein member 3 precursor (SM-11044
DE binding protein) (EP70-p-iso).
GN TM9SF3 OR SMBP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Warner S.J., Lomax M.I.;
RT "Evolution of the TM9 super family of membrane spanning proteins.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RA Gu Y., Peng Y., Li Y., Fu S., Gu J., Gu W., Jiang C., Yu Y., Han Z.,
RA Wang Y., Chen Z., Fu G.;
```



```
QY 151 -----ENSYGVGGDPHTYIGALLIGN-----NIC-GLLITVVKVIGVYFLNDPEP 195
Db 238 PTPVSALLHSSTMVAGVFLIRFPLMENNKTIOSLTLCICLGAITTLTAICALTQNDIK 297
QY 196 KLVA-----IVYRGIS---LVILLVCAIALF-----FITKQDFYHHHOKMEIRE 238
Db 298 KIIAFSTSQLGLMTIVTIGINOPYLAFLHICHTHAFFKAMLFMCSGSIISHLNDE-ODIRK 356
QY 239 KAETDRPSPSLMTFTTCYQGL-----FNWVFCFAVLT 273
Db 357 MGLLENAMP---FTTSLIISGLALTGPFLTGFYSKDLIETANTSTNAW---ALLMT 410
QY 274 IFPVNMTV--TTR-----GDSGF--LNKIMSENDEIYTLTSLFVFNFAAIGSIVAS 322
Db 411 LIATSLTAVSTRIFFALLGPRLPLTSINENNPFINSIKRLIGSIFA--GFFISN 468
QY 323 KIHWT-----PRYLFAILRALFPFFPFC-----NRYVOTRAYPVFFEST 365
Db 469 NI-YPTTVPEMTPTYMK-----LTALAVTILGFTLALSLMTHNKLKLE-----HST 515
QY 366 DIFVIGGIAMSFHGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLTGLW----- 420
Db 516 NVF-----KFSN-----LGIYPTTM--HRLPPLANLSMSQKSASLLDLSIWLENIL 560
QY 421 PVVIEHFVDKPSIL 434
Db 561 PKSISQFMQKTSIL 574

RESULT 14
TCR2_BACSU STANDARD; PRT; 459 AA.
AC P14512;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Tetracycline resistance protein.
GN TET.
OS Bacillus subtilis.
OC Plasmid pNS1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Noruchi N., Aoki T., Sasatsu M., Kono M., Shishido K., Ando T.;
RT "Determination of the complete nucleotide sequence of pNS1, a
RT staphylococcal tetracycline-resistance plasmid propagated in Bacillus
RT subtilis."
RL FEMS Microbiol. Lett. 37:283-288(1986).
CC -!- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE
CC ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTI-PORTER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16217; AAA19179.1; -
DR PIR; S42238; S42238.
DR InterPro; IPR001411; TCR_TetB.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_ttr; 1.
DR PRINTS; PK01036; TCR_TETB.
```

```
KW Antibiotic resistance; Transmembrane; Transport; Symport; Plasmid.
FT TRANSMEM 12 33 POTENTIAL.
FT TRANSMEM 81 100 POTENTIAL.
FT TRANSMEM 111 129 POTENTIAL.
FT TRANSMEM 140 162 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 223 240 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 346 365 POTENTIAL.
FT TRANSMEM 432 451 POTENTIAL.
SQ SEQUENCE 459 AA; 50695 MW; FE53C55535372B41 CRC64;

Query Match 4.9%; Score 113; DB 1; Length 459;
Best Local Similarity 18.1%; Pred. No. 0.59;
Matches 82; Conservative 79; Mismatches 133; Indels 160; Gaps 20;

QY 22 YIVVYLVLVGVGLLPWNMFITIAPEYVYVWFKPDGVETWYKSEPMGSLTIGSOLPN 81
Db 13 YSVLFWLCLLSFSSVLN--EMVYNVSLPDIAHFNTTPOITWVWYAYMLTFSIGTAVYG 70
QY 82 ASINVENL-FLIAGPLIYRVFAPVCFNIVNLTIIILVIVLEPTEDSDMSWFFWTLGMA 140
Db 71 KLSVDYINIKKLLIIG-----ISLCLGLSLIAFI-----GHNHFFILIFG-- 109
QY 141 TSINFSNGLYENSVYGVG-GDFPHT-----YIGALL-----IGNNI 175
Db 110 -----RLVQGVGSAAPSLMVVARNITRKKOGKAFGTGSIVALGEGLGPSI 158
QY 176 CGL-----LITVVKIGVTVFLNDEPKLVAIVYFGISL-----VILLVC 213
Db 159 GGIHAIYHWSYLLILPMTIVTI-----PELIKVMVPGKSKNTKNDIVIGVILMSI 209
QY 214 AIALFFITKQDFYHHQKMEIREKAETDRPSPSILMTFTTCYCGQLFNVMVFCFAVLT 273
Db 210 SIICFML-----FTTNVNWTFILFTI----- 231
QY 274 IFPVNMTVTTRGDSGLNKMSEN-DEIYTLTSLFVFNFAAIGSIVASKIHWTPRYL 332
Db 232 FFVIFIKHISRVSNPFINPKLGKIPFMLGSLFSGGLIFSIVAGFSMWVPMY--KTIYHV 289
QY 333 KPAILRALFIP-----FFFCNVYRVOTRAYPVFFESTDIFVIGGIAMSFHGYLSAL 385
Db 290 NVATTGNSVIFPGTMSVIVFGYGGFLVDRKG-----SLFVILGSLIS-----ISFL 338
QY 386 AMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGL 419
Db 339 TIAF-----FVEFSMWLT--TFMFIVMGGL 362

RESULT 15
YDNK_LACLC STANDARD; PRT; 614 AA.
ID YDNK_LACLC
AC P42377;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 70.0 kDa protein in dnaK 3'region (ORF4).
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MG1363;
RX MEDLINE=94172317; PubMed=8126443;
RA Eaton T.J., Shearman C.A., Gasson M.J.;
RT "Cloning and sequence analysis of the dnaK gene region of Lactococcus
RT lactis subsp. lactis."
RL J. Gen. Microbiol. 139:3253-3263(1993).
```

```
CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76642; CAA54090.1; -.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 69869 MW; B1146136D09B89D9 CRC64;

Query Match
Best Local Similarity 4.9%; Score 113; DB 1; Length 614;
Matches 98; Conservative 66; Mismatches 149; Indels 196; Gaps 26;

QY 21 KYNIVYWLVLVGVGLLPWNMFITIA---PEYVYNY----- 54
DB 16 KYNFEFYIIL-----FITLATFPYRGNFHAGNDFAFNARVNSTISALKDG 64

QY 55 -----WFKPDGVE-----TWYSEFGMSLIGSOLPNASINVENFLIIAGPLIYRVFAPVC 106
DB 65 QVIOQDPNALSGFGYAW--NEFYG-----PLPTYFISVIK-FIVKWSLSFSLSFYSLC 115

QY 107 FNVNLTILILIVLEPTDSMSWFFWVTGMAISINFSNGLYENSYGVGGDPHTYI 166
DB 116 LFISGIFTFNSSFULKDHNTS-KLFGLLAVALFT---FSNSTYINLYYANPSQLALL 171

QY 167 GALLI--GNN-----ICG-----LITVVVKIGVT 188
DB 172 FVILLFWGMNKNKRSFAFLVAFVAGAGLPLSHVTVTICTLPFVLLYLLLIKKG-- 229

QY 189 YFLNDEPKLVAIVFGISLVILLVCAIALFF-----ITKODFYH----- 227
DB 230 -NLKENIKIIGLGLSVTSAI---GLSAFFLPLENLKSGIYNVNSDFSRSPGWNNI 284

QY 228 -YHHOKGMEIREKAETDRPSPILWTF-----TNC-YGQLENVWFCEPAVT 271
DB 285 AYFOGKWEPLKIEFSYKFPSSLFVFLVFIFISLINFKNKTNAKYSLIFS--CFSLV 341

QY 272 LTIFPVMVTTRGDSGFLNKIMSENDEIYTLTS---FLVFNLFAAIG-----SIVAS 322
DB 342 LVLMQLPI-----FPWKIFS---IFTIVQDPARFSTLGLFSLSLVLILPILLD 388

QY 323 KIHWPTRPYLFAILRALFIPFFFCYRVQTRAYPVF-----FESTDIF 368
DB 389 KISGKTSYLYLTGLLVIFSLGFAEFNR-RIOKGSQPLFASQAQSLLNKTPFNMYMENPDSI 447

QY 369 VIGGIAMSFSGYLSALAMGYTPNVVPSH 397
DB 448 AIG-----EYLPQVIGSH 460
```

Search completed: July 23, 2002, 16:43:23  
Job time: 496 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 15:41:32 ; Search time 44.23 Seconds  
(without alignments)  
942.862 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVIINRSNTYAVEQAFPRDKYNIYVWLVGVGLLPWNMFIITAIPEYYVNYWFKPDG 434  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Keeping first 45 summaries

Database :

PIR\_71: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2265	98.7	434	2	T16080
2	716.5	31.2	461	2	T28006
3	706.5	30.8	450	2	T23528
4	574.5	25.0	800	2	T23247
5	313.5	13.7	428	2	T01489
6	249.5	10.9	408	2	C86156
7	214	9.3	326	2	JC4196
8	209.5	9.1	327	2	JC4195
9	196.5	8.6	513	2	T21887
10	182.5	8.0	517	2	S36712
11	180	7.8	143	2	T22164
12	144	6.3	418	2	D85064
13	138.5	6.0	418	2	E85064
14	134	5.8	432	2	H64383
15	130	5.7	425	2	E97095
16	128.5	5.6	382	2	G96641
17	127.5	5.6	581	2	E90539
18	126	5.5	418	2	F83986
19	125.5	5.5	674	2	T21217
20	121	5.3	492	2	E90373
21	120	5.2	330	2	C95844
22	118	5.1	498	2	F89861
23	117.5	5.1	506	1	D64048
24	117.5	5.1	559	1	S62503
25	117	5.1	463	2	F90285
26	115.5	5.0	413	2	E95041
27	115	5.0	456	2	A90013
28	115	5.0	461	2	B83601
29	114.5	5.0	449	2	C75053

ALIGNMENTS

RESULT 1

T16080

hypothetical protein F16H11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Jun-2000

C:Accession: T16080

Submitted to the EMBL Data Library, April 1996

A:Accession: The sequence of C. elegans cosmid F16H11.

A:Reference number: Z18458

A:Accession: T16080

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-434 <WUX>

A:Cross-references: EMBL:U55376; NID:g1280130; PID:g1280131; PIDN:AAA98003.1; GSPDB:G

A:Experimental source: strain Bristol N2; clone F16H11

C:Genetics:

A:Gene: CESP:F16H11.3

A:Map position: X

A:Introns: 49/3; 94/2; 190/1; 256/3; 326/3; 392/2

C:Superfamily: Caenorhabditis elegans hypothetical protein ZR809.4

Query Match 98.7%; Score 2265; DB 2; Length 434;  
Best Local Similarity 99.1%; Pred. No. 2e-172;  
Matches 430; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MVIINRSNTYAVEQAFPRDKYNIYVWLVGVGLLPWNMFIITAIPEYYVNYWFKPDG 60

Db 1 MVIINRSNTYAVEQAFPRDKYNIYVWLVGVGLLPWNMFIITAIPEYYVNYWFKPDG 60

Qy 61 VETWYSKFMGSLTIGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILVLI 120

Db 61 VETWYSKFMGSLTIGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILVLI 120

Qy 121 VLEPTEDSMNSFFWVTLGMATSNFNGLYSVGVGDPPHTYIGALLIGNNICGLLI 180

Db 121 VLEPTEDSMNSFFWVTLGMATSNFNGLYSVGVGDPPHTYIGALLIGNNICGLLI 180

Qy 181 TVWIKGVTVFLNDEPKLVAIVFVGLSVLLVCAITALEFFITKQDFYHHQGMREKA 240

Db 181 TVWIKGVTVFLNDEPKLVAIVFVGLSVLLVCAITALEFFITKQDFYHHQGMREKA 240

Qy 241 ETDPRSPSILWTTFTNCYQGLFNWFCFAVLTITPPVMMVTVTTRGDSGFLNKIMSENDEI 300

Db 241 ETDPRSPSILWTTFTNCYQGLFNWFCFAVLTITPPVMMVTVTTRGDSGFLNKIMSENDEI 300

Qy 301 YTLTSTFLVFNLAIGSVASKIHWPTPRYLKFAILLRALFIPFEFCNRYVQTRAYPV 360

Db 301 YTLTSTFLVFNLAIGSVASKIHWPTPRYLKFAILLRALFIPFEFCNRYVQTRAYPV 360

Qy 361 FFESTDIEFVIGIAMSFSGVLSALMGYTPNVVPSYRSFAAQLSVCTLMVGLLTGGLW 420







Db 273 FP-----GFAENLKSQSLQSWYPILLITVYNISDFVGKSLFALYLWQSIKAT 321  
Qy 334 FAIIILRALFIPEFFPCNRYQVTRAPVFFESDIFVIGGIAMSFSGHYLSALAMGYTPNV 393  
Db 322 WACIVRLLFYPLFSAC-----LRGPKWLREVPVVLTFMGLTNGLYLTVLMIMAPKT 375  
Qy 394 VPSHYSRFAAQLSVCTLMVGLLTGG-----LW 420  
Db 376 VHASEAELAFMVVFLGLGLVCGSVIGLW 406  
  
RESULT 7  
JC4195  
36k hydrophobic nucleolar protein - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 21-Jul-2000  
R:Accession: JC4195  
R:Williams, J.B.; Lanahan, A.A.  
A:Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein  
A:Reference number: JC4195; MUID:95367016  
A:Accession: JC4195  
A:Molecule type: mRNA  
A:Residues: 1-326 <WIL>  
A:Cross-references: EMBL:X86681; NID:951266; PIDN:CAA60380.1; PID:951267  
A:Experimental source: heart  
C:Comment: This protein has a role in the growth response and participates in the proliferation  
C:Genetics:  
A:Gene: dex12  
A:Start codon: AUG  
C:Keywords: nucleoprotein; phosphoprotein  
F:54/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F:107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 9.3%; Score 214; DB 2; Length 326;  
Best Local Similarity 22.0%; Pred. No. 1.2e-09;  
Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;  
  
Qy 143 INFNSGLYENSVYGVGGDPHTYIGALLIGNNICGLLITVVKI-----GVTVFLNDEPKL 197  
Db 7 INSFSAVLQGSFLGQLTMPSTYSLFSLGQGLAGIFAALAMLLSMASGV-----DAET 60  
  
Qy 198 VAIVYF---GISVLVLCALFFITKQDFY---HYHQKGMEIREKA----- 240  
Db 61 SALGYFITPYVYILMSIVCYLSPLHLKFARYLLANKSSQAQAELETAEKALLQSDENGIP 120  
Qy 241 -----ETDRSPSILWTTFTNCYQQLPNVWFCFAVTLTIPPV 277  
Db 121 SSPQKVALTLDLLEKEPESEPEPKPKPSVFTVFOKIMLTALCLVLVFTVTLVSFPA 180  
Qy 278 MMTVTTRGDSGFLNKIMSENDEIYLLTSFLVFNLFRAIGSIVASKIHWP--TPRYLKFA 335  
Db 181 ITAMVTSSTS-----PKWQSFNPICCFLLFNIMDMWLGRLSTSYFLWPDSDRLPL 234  
Qy 336 IILRALFIPEFFPCNRYQVTRAPVFFESDIFVIGGIAMSFSGHYLSALAMGYTPNVVP 395  
Db 235 VCLRFLFVPLFMCHVPQSR-LPILFQDAYFITFMLFAVSNGLYLSLTMCLAPROVL 293  
Qy 396 SHYSRFAAQLSVCTLMVGLLTG 417  
Db 294 PHEREVAGALMTFFLALGLSCG 315  
  
RESULT 8  
JC4195  
36k hydrophobic nucleolar protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 05-Nov-1999  
R:Accession: JC4195  
R:Williams, J.B.; Lanahan, A.A.  
A:Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein

A:Reference number: JC4195; MUID:95367016  
A:Accession: JC4195  
A:Molecule type: mRNA  
A:Residues: 1-327 <WIL>  
A:Cross-references: EMBL:X86682; NID:951302; PIDN:CAA60381.1; PID:951303  
C:Comment: This protein has a role in the growth response and participates in the proliferation  
C:Genetics:  
A:Gene: dex12  
A:Start codon: AUG  
C:Keywords: nucleoprotein; phosphoprotein  
F:82-159/Domain: hydrophilic #status predicted  
F:54/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F:107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 9.1%; Score 209.5; DB 2; Length 327;  
Best Local Similarity 22.4%; Pred. No. 2.8e-09;  
Matches 76; Conservative 49; Mismatches 134; Indels 81; Gaps 12;  
  
Qy 131 WPFVWTLGMAITSINFNSGLYENSVYGVGGDPHTYIGALLIGNNICGLLITVVKI----- 185  
Db 5 WF-----INSFCAVLQGSFLGQLTMPSTYSLFSLGQGLAGIFAALAMLLSLAS 54  
Qy 186 GVTYFLNDEPKLVAIVYF---GISVLVLCALFFITKQDFY---FFI-----TK 222  
Db 55 GV-----DAOTSALGYFITPCVIGILLSIVCYLSPLHLKFARYLLTEKLSQAPTQLETK 108  
Qy 223 QDFYHYHQKGMEIR-----EKAETDRSPSILWTTFTNCYQQLPNVWF 266  
Db 109 AELLQADEKNGVPISPQOASPTLDLDPEKEPEPEPKPKPSVFTVFOKIMLTALCLVL 168  
Qy 267 CFAVTLTIPPVN--MTVTTRGDSG---FLNKIMSENDEIYLLTSFLVFNLFRAIGSIV 320  
Db 169 VFTVTLVSFEPAITAMVTTSSNSPGKWLFFNP-----CCFLLFNVMDWLGRL 217  
Qy 321 ASKIHP---TPRYLKFAIILRALFIPEFFPCNRYQVTRAPVFFESTDIFVIGGIAMSF 377  
Db 218 TSYFLWPDSDRLPLVCLVFLFVPLFMCHVPQSR-LPILFQDAYFITFMLFAV 276  
Qy 378 SHGYLSALAMGYTPNVSPSHYSRFAAQLSVCTLMVGLLTG 417  
Db 277 SNGYLVSLTMCLAPROVLPHEREVAGALMTFFLALGLSCG 316  
  
RESULT 9  
T21887  
hypochemical protein F36H2.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21887  
R:Steward, C.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19483  
A:Accession: T21887  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-513 <WIL>  
A:Cross-references: EMBL:Z81078; PIDN:CAB03075.1; GSPDB:GNO0019; CESP:F36H2.2  
A:Experimental source: clone F36H2  
C:Genetics:  
A:Gene: CESP:F36H2.2  
A:Map position: 1  
A:Introns: 28/3; 56/1; 93/2; 163/3; 277/3; 314/1; 341/3; 384/2; 407/2; 458/2; 480/3  
  
Query Match 8.6%; Score 196.5; DB 2; Length 513;  
Best Local Similarity 21.4%; Pred. No. 5e-08;  
Matches 95; Conservative 77; Mismatches 159; Indels 113; Gaps 21;  
  
Qy 31 LVGFGVLPWNFFITAEYYVNYWFKPDGV-----ETWYSKPEMGSLITGSOLP 80  
Db 11 IVSISMFNFNAFMN-AHEY---FYIKLRNVNTDNDPVTNTDTWFIKR-----QIRDOVP 61



Query Match 6.3%; Score 144; DB 2; Length 418;  
Best Local Similarity 20.6%; Pred. No. 0.00059;  
Matches 96; Conservative 76; Mismatches 159; Indels 134; Gaps 23;

QY 18 PRKYNIWVLI---LVGQVLLPWNMFITIAPEYV-VNYWFKPDGVETWYSKEF-MGS 72  
DB 9 PPELQKQYQAMVVCILGIGSLVSNMMLTIADYYVKKVFPDYHPSRVLTFLVYQPEALGT 68  
QY 73 LTI----GSQLPNASINVENFLIIAGPLIYRVFAPVCNIVNLITILIVIVLEPTEDS 128  
DB 69 ILTAYHESKINTRKRNLIYIL-----FTISTELLIVLD----- 103  
QY 129 MSFEFWTLGMAT-----SINFSNGLYENSVMG--VGG-----DPPHTYIG 167  
DB 104 -----LATKGGGIGPYIGLCVAVNASVGLADATVOGGMIGDLSLPCPELVOSFMG 153  
QY 168 ALLIGNNICGLLTVTKVIGVYFLNDEPKLVAIVYFGISVILLVCAIALFFI-TKQDFY 226  
DB 154 GLAVSGALTSALRLITKAAFEK-TNDGPRKGAAMFLAISTCIELLCVFLYAYVFPKLPV 212  
QY 227 HYHQKMEIREK-----AETDRPSPSILWTTTNCYQQLFNWV 265  
DB 213 KYRRAASEGSKTVSADLAAAGIQNSDLTDDDSKNQRLSNKELLIQNIDYAVNLFYI 272  
QY 266 FCFPAVLTITFPVMMVTTRGDSGLFNKINSEN--DEIYTLTSLFLVNLFAAIGSIVASK 323  
DB 273 VC---TLSIFP-----GFLYENTGOHGLGDWYALVL-VAMYNCDLVGR----- 312  
QY 324 IHWPTRYLK-----FAILRLALFIPFFFCNRYRVOTRAYPVFFESTDIFVIGGIAM 375  
DB 313 -YTPLVKWLKIENRKLTITIAVLSRYLLIPAFYF-TAKYGDQGMIML-----ISVLG--- 362  
QY 376 SPFHGYLSALAMGYTNVPVSHYSRFAAQLSVCTLMVGLLTGGLW 420  
DB 363 -LTNGHLTVCM---TIAPKGY-KGPEQNALGNLLVIFLLGGIF 401

RESULT 13  
E85064  
hypoetical protein AT4g05130 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: E85064  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: E85064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-418 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267272; PIDN:CAB81055.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g05130  
A:Map position: 4

Query Match 6.0%; Score 138.5; DB 2; Length 418;  
Best Local Similarity 21.0%; Pred. No. 0.0016;  
Matches 95; Conservative 73; Mismatches 174; Indels 111; Gaps 23;

QY 21 KYNIVWVILVFGVLLPWNMFITIAPEYV-VNYWFKPDGVETWY-SKEFMGSLRIGSQ 78  
DB 16 KQAMVVCIL-GIGSLFWSNMLTIADYYVFPDYHPSRVFTLLYQPIALGTIMI--- 71  
QY 79 LPNASINVENFLIIAGPLIYRVFAPVCNIVNLITILIVIVLEPTEDSMSEFWWTGL 138  
DB 72 LAYRESKISTRILTYILF-----TISTELLIVLDLT-----TKG 108  
QY 139 -----MATSNFSNGLYENSVMY-GVGDDFP---HTVIGALLIGNNICGLLITVVKIG 186  
DB 109 HGGIGHYIVLCTIVASFGLDATVKGGLVGDLSLPCPELIQSYMGSMAGALTSVLR- 167

QY 187 VTVEL---NDEPKLVAIVVFGISVILLVCAIALFFI-TKQDFYHYHHQK----- 233  
DB 168 ITRAAAEKSNNSURKAMFLAISTFIELLCVLIVAYVFKPLVIVKYRKAASEGSKTV 227  
QY 234 -----MEIREKAETD-----RPSPSILWTTTNCYQQLFNWFCFAVTLTIFPVM 278  
DB 228 VADLAAAGIQNSLSDSDSKNQMLRKKELLQN-----IDHAVNLFLIVLTLISIFF-- 280  
QY 279 MTVTTRGDSGLFNKINSEN--DEIYTL--TSLVNLNFAAIGSIVASKIHW---PTPRY 331  
DB 281 -----GFLYENTGOHGLGDWYALIVATYFNWDLFGRYAPLV---KWLKLENRKA 327  
QY 332 LKFAIILRALFIPFFFCNRYRVOTRAYPVFFESTDIFVIGGIAMSPFSGHYSALAMGYTP 391  
DB 328 LTATVTRVFLVPAFYF-TAKYGDKGMMIMLV-----IIGLTTGHLTVCIIMTIAP 377  
QY 392 NVVPVSHYSRFAAQLSVCTLM---VGLLTGGLW 420  
DB 378 NGYKGPKEKNALGNLLVVFIILGGAUVGSLGLWLW 410

RESULT 14  
H64383  
Nat transporter - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 31-Dec-2000  
C:Accession: H64383  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996  
A:Authors: kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc A:Reference number: A64300; MUID:96337999  
A:Accession: H64383  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-432 <BUL>  
A:Cross-references: GB:U67514; GB:L77117; NID:g2826304; PIDN:AAB98666.1; PID:g1591386  
C:Genetics:  
A:Map position: REV598993-597695  
C:Superfamily: probable transporter MJ0672

Query Match 5.8%; Score 134; DB 2; Length 432;  
Best Local Similarity 19.1%; Pred. No. 0.0038;  
Matches 82; Conservative 80; Mismatches 133; Indels 134; Gaps 22;

QY 66 SKEFMG-----SLTIGSQLPNASINVENFLIIAGPLIYRVFAPVCNIVNLITILV 119  
DB 4 SKEFGLGITASLIFGSSLPDIYKGI--VILIVACLWFFELLPLPVTSLAIPIMAVFL 61  
QY 120 IVLEPTEDSMWF-----FWVTLG---MATSNFSN--GLYENSVMYGVGGDFPHF----- 164  
DB 62 GIFN-LKEALTYPFAHPITFIIFLFGGFLAQAALKNHNLDKFIAYKLLNYGDKFTCTCFMLF 120  
QY 165 --YIGALLIGNNICGLLITVVKIGVTV-----FLNDEPKL 197  
DB 121 SAYFLSMWISNTSATLILPAILGGLLHKTKGRDLRDFLLGVAYSASIGGIATIGSGPNA 180  
QY 198 VAIVY-----FGISVILLVCAIALF-----FITKQDFYHYHHQKMEIREKA 240  
DB 181 IASSYLDYGFSSWFKVGFPPISLLFLLCILCTLYIYFKWIPKED-----IAIQARM 231  
QY 241 ETRDRPSPI-----LWTTTNCYQQLFNWFCFAVTLTIFPVMVMTTRGDSGLFN 291  
DB 232 ELSRNAYKLLVIFVLIASLW-IISDYLSEIFNVQY-FDSVIAIFAILL-----FVF 281  
QY 292 KIMSEND---EITYTL-----TSFLVNLFAAIGSIVASKIHWPTPR 330  
DB 282 NLVEVNDFKKIDMGWTLILFGALCLGVIVKSGANTFLSEKLIAILGNL-----TPI 333

Search completed: July 23, 2002, 16:36:11  
Job time: 3279 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 15:37:31 ; Search time 24.8 seconds  
(without alignments)  
427.448 Million cell updates/sec

Title: US-08-816-011f-63

Perfect score: 2294

Sequence: 1 MVIIINRNTYAVEQEAFFRD.....LTGGLWPVVEHFVDKPSIL 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 25 summaries

Database :

- 1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	343.5	15.0	456	4	US-09-058-389A-4
2	326.5	14.2	456	4	US-09-058-389A-2
3	214	9.3	326	4	US-09-058-389A-3
4	181	7.9	247	4	US-09-058-389A-8
5	111.5	4.9	398	1	US-08-097-938-8
6	111.5	4.9	398	1	US-08-476-000-4
7	111.5	4.9	398	1	US-08-472-840-4
8	111.5	4.9	398	2	US-08-476-976-4
9	111.5	4.9	398	3	US-08-474-410-4
10	111.5	4.9	398	4	US-08-486-673B-4
11	109.5	4.8	397	4	US-08-486-673B-63
12	108.5	4.7	398	1	US-08-097-938-6
13	108.5	4.7	398	1	US-08-476-000-6
14	108.5	4.7	398	1	US-08-472-840-6
15	108.5	4.7	398	2	US-08-476-976-6
16	108.5	4.7	398	4	US-08-474-410-6
17	108.5	4.7	398	4	US-08-486-673B-6
18	106.5	4.6	395	1	US-08-097-938-2
19	106.5	4.6	395	1	US-08-097-938-5
20	106.5	4.6	395	1	US-08-476-000-2
21	106.5	4.6	395	1	US-08-476-000-5
22	106.5	4.6	395	1	US-08-472-840-2
23	106.5	4.6	395	1	US-08-472-840-5
24	106.5	4.6	395	2	US-08-476-976-2
25	106.5	4.6	395	2	US-08-476-976-5
26	106.5	4.6	395	3	US-08-474-410-2
27	106.5	4.6	395	3	US-08-474-410-5

28	106.5	4.6	395	4	US-08-486-673B-2	Sequence 2, Appli
29	106.5	4.6	395	4	US-08-486-673B-5	Sequence 5, Appli
30	106.5	4.6	399	1	US-08-476-000-61	Sequence 61, Appl
31	106.5	4.6	399	1	US-08-472-840-61	Sequence 61, Appl
32	106.5	4.6	399	2	US-08-476-976-61	Sequence 61, Appl
33	106.5	4.6	399	3	US-08-474-410-61	Sequence 61, Appl
34	106.5	4.6	399	4	US-08-486-673B-61	Sequence 61, Appl
35	105.5	4.6	397	1	US-08-476-000-63	Sequence 63, Appl
36	105.5	4.6	397	1	US-08-472-840-63	Sequence 63, Appl
37	105.5	4.6	397	2	US-08-476-976-63	Sequence 63, Appl
38	105.5	4.6	397	3	US-08-474-410-63	Sequence 63, Appl
39	101	4.4	420	2	US-08-466-103A-2	Sequence 2, Appli
40	96	4.2	473	1	US-08-597-236-13	Sequence 13, Appl
41	96	4.2	473	1	US-08-746-682A-13	Sequence 13, Appl
42	95.5	4.2	336	3	US-08-749-816-2	Sequence 2, Appli
43	95.5	4.2	336	4	US-09-144-914-2	Sequence 2, Appli
44	94.5	4.1	436	6	5432081-10	Patent No. 5432081
45	94	4.1	418	4	US-09-030-267-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-058-389A-4  
; Sequence 4, Application US/09058389A  
; Patent No. 6130065  
; GENERAL INFORMATION:  
; APPLICANT: Belt, Judith A.  
; APPLICANT: Crawford, Charles R.  
; APPLICANT: Patel, Divyen  
; TITLE OF INVENTION: A NITROBENZYLIMBECAPTOPYRINERIBOSIDE  
; TITLE OF INVENTION: (NBMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT  
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF  
; TITLE OF INVENTION: USE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
; STREET: Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/058,389A  
; FILING DATE: April 9, 1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 1340-1-013N  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 456 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; DESCRIPTION: hENT1  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-09-058-389A-4









```

; Sequence 4, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,840
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-472-840-4

Query Match          4.9%; Score 111.5; DB 1; Length 398;
Best Local Similarity 18.5%; Pred. No. 0.0045;
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

QY   61 VETWYS-KEPGSLTIGSQLPNASINVFNFLIIAGPLIYRVFAPVCFNIVNLTIILYL 119
    ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   56 VETVFVSDEFSASVLTKG-----LTTVELPIVTIV-----F 87

QY   120 IVLEPTEDSMSEFF-----WWTLGMATSNFNSGLYENS 153
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   88 VVGLPNGMALWFLEFRKKKHPAVIYMANLAADLLSVIFPKLIAYHIGNNWYGEA 147

QY   154 YVGGDFPHTYIGALLIGNICGLITVVYKIGVTYFLNDP----KLVAIVFGISLVI 209
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   148 LCNV-----LIG-FFYGNMYCSILEFTCLSVQRVWIYNPMGHSHKRKAIAIGISLAI 199

QY   210 -LIVCAIAL-FFTICKODFYHHQKGMEIREKAETDRPSLSILWTFTTCNY----- 258
    || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   200 WLILLVTPLYVVKOTIF-----IPAL---NITTCHDVLPEQLLV 237

QY   259 GOLFNWFCFAVLTITPPVMVTTRGDSGFL-----NKIMSENDE-----IYT 302
    |:|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   238 GDMFNFLSLAIGVLFPAELTA-----SAYVLMIRLRSSAMDENSEKKRRRAIKLIVT 292

QY   303 LLTSFLVNFLAAIGSIVASKIHWPTRYLKFAILLRALFIPEFFCNYEVQTRAPVVEF 362
    :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db   293 VLAWYLI--CF-----TPSNL-----LLVHYFLIKSGQGRHVVALYI 328

QY   363 ESTDIFVIGGIAMSFHGYSALAMGYTPNVWPHYSRFAAQSLV 407

```













GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 15:36:31 ; Search time 62.22 Seconds  
(without alignments)  
774.767 Million cell updates/sec

Title: US-08-816-011f-63

Perfect score: 2294

Sequence: 1 MVIINRNTYAVEQEAFPRD.....LTGGLWPVVIEHVDKPSII 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2294	100.0	434	22	AAU07630
2	2290	99.8	434	17	AAR92315
3	645	28.1	458	22	ABB64784
4	372.5	16.2	475	21	AA182286
5	343.5	15.0	456	21	AAW69555
6	343.5	15.0	456	21	AAB15520
7	327.5	14.3	475	21	AA192285
8	327.5	14.3	475	22	AA192285
9	326.5	14.2	456	19	AAW67487
10	326.5	14.2	456	19	AAW69558
11	326.5	14.2	456	21	AAB15517

12	326.5	14.2	475	21	AA199363	Human PRO1380 (UNQ
13	326.5	14.2	475	22	AAU29153	Human PRO polypept
14	326.5	14.2	475	22	AAB66112	Protein of the inv
15	320.5	14.0	456	19	AAW69557	Rat equilibrative
16	301	13.1	457	19	AAW69556	Rat equilibrative
17	272.5	11.9	373	19	AAW64550	Human fibrosarcoma
18	261.5	11.4	397	22	AAW63550	Human protein sequ
19	221.5	9.7	586	22	ABW65393	Drosophila melanog
20	217.5	9.5	404	22	ABW68480	Drosophila melanog
21	214	9.3	326	21	AAB15519	Human HNP36 protei
22	181	7.9	247	19	AAW83966	Alternatively splic
23	181	7.9	247	21	AAB15518	Human NBMPR-iENTP
24	180	7.8	314	20	AAW89005	Polypeptide fragme
25	180	7.8	314	22	ABW51083	Human secreted pro
26	179.5	7.8	423	22	AAB88377	Human membrane or
27	172	7.5	248	22	AAW85230	Human immune/haema
28	125	5.4	115	22	AAW90509	Human immune/haema
29	124	5.4	521	22	ABW62910	Drosophila melanog
30	119.5	5.2	383	22	AAG82466	S. epidermidis ope
31	117	5.1	866	22	ABW64833	Drosophila melanog
32	114.5	5.0	450	22	AAW96525	Putative P. abyssi
33	114.5	5.0	527	21	AAW76126	Human secreted pro
34	113.5	4.9	329	22	AAW93524	Human polypeptide,
35	113.5	4.9	329	22	AAW92687	Human protein sequ
36	113.5	4.9	360	22	AAW75546	Human secreted pro
37	113.5	4.9	530	22	AAW75598	Human secreted pro
38	113.5	4.9	545	22	AAW94667	Human protein sequ
39	113.5	4.9	579	21	AAW82459	Human SM-11044-bin
40	113.5	4.9	582	21	AAW82459	Human SM-11044-bin
41	111.5	4.9	397	21	AAB35641	Human PAR-2 protei
42	111.5	4.9	398	16	AAW66921	Human C140 recepto
43	111.5	4.9	398	17	AAW01953	Human C140 recepto
44	111.5	4.9	545	21	AAW94910	Human secreted pro
45	111	4.8	439	19	AAW61371	Non-adrenergic SM

ALIGNMENTS

RESULT 1  
AAU07630  
ID AAU07630 standard; Protein: 434 AA.  
XX  
AC AAU07630;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Caenorhabditis elegans potassium ion channel CORK protein.  
XX  
KW Transmembrane potassium ion channel protein; inward potassium flux;  
KW pest control; membrane potential; pesticide; antihelminthic; nematode;  
KW insect; CORK.  
XX  
OS Caenorhabditis elegans.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 7 /note= "N-glycosylated"  
FT Domain 26...39 /note= "Putative pore-forming H5 domain #1"  
FT Modified-site 83 /note= "N-glycosylated"  
FT Modified-site 146 /note= "N-glycosylated"  
FT Domain 150...162 /note= "Putative pore-forming H5 domain #2"  
XX WO200161006-A2.  
XX 23-AUG-2001.  
XX 14-FEB-2001; 2001WO-US04680.

```
PR 15-FEB-2000; 2000US-0503849.
XX (BADI ) BASF CORP.
XX Pausch MH;
XX WPI; 2001-536570/59.
DR N-PSDB; AAS12162.
XX
PT New polypeptide, a mutant potassium ion channel protein for improving
PT inward potassium flux under acidic conditions
PS Example 14; Fig 9; 131pp; English.
XX
CC The invention relates to a mutant potassium ion channel protein, having
CC four membrane spanning domains and two pore forming domains, comprising a
CC mutation at the second pore forming domain. The expression of the mutant
CC protein in a cell confers improved inward potassium flux and the ability
CC to grow in the presence of potassium. Mutant proteins and their
CC corresponding polynucleotide sequences can therefore be used to improve
CC inward potassium flux into cells under acidic conditions by modulating
CC the membrane potential using therapeutic agents. The sequences may be
CC used to develop agonists and antagonists of potassium channel proteins in
CC order to control pests such as nematodes and insects. This sequence
CC represents a Caenorhabditis elegans transmembrane potassium ion channel
CC protein, CORK.
XX
SQ Sequence 434 AA;
```

Query Match 100.0%; Score 2294; DB 22; Length 434;  
Best Local Similarity 100.0%; Pred. No. 8.7e-228;  
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVIINRSNTYAVEQAPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDG 60  
Db 1 mviinrsntyaveqaprdkyniyvwlvilvfgvllpwnmfitiapeyyvnywfkpdg 60  
Qy 61 VETWYSKEFMGSLTIGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILILVI 120  
Db 61 vetwyskefmgsltigsolpnasinvnflfliagpllyrvfapvcfnvltlililvi 120  
Qy 121 VLEPTEDSMFVFWTLGMATSFNSGLYNSVYGVGDFPHTYIGALLIGNNICGLLI 180  
Db 121 vleptedsmsfwtlgmatsfnsglyensvygvvgdfphtyigallignnicglli 180  
Qy 181 TVWKIGTVYFLNDEPKLVAIVFGISVLILLVCATALFFITKQDPIYHHQGMIREKA 240  
Db 181 tvwkigtvyflndepklvaivfgisvlillvcatalffitkqdfyhhqgmireka 240  
Qy 241 ETDPRSPSLMTFTNCGYGFVWFCFAVTLTIPVMMVTTRGDSGFLNKMSENDEI 300  
Db 241 edrpspslmtftncgygfwnwfcfavtltipvmmtvttrgdsfglnkmsendei 300  
Qy 301 YTLTSLFVLNFAAIGSIVASKIHWPTRPYLKFAILRALFIPFFFCNTRVQTRAYPV 360  
Db 301 ytltsflvlnfaaigsivaskihwptprylkfaillralfipfffcntrvqtraypv 360  
Qy 361 PFESTDIFVIGGIANSFSGYLSALAMGYTNVPVSHSRFAAQLSVCTLMVGLLTGGLW 420  
Db 361 pfestdifvigiamsfshgylsalamgytnvpvshysrfaaqlsvctlmvglgtglw 420  
Qy 421 PVVIEHFVDKPSIL 434  
Db 421 pvviehfvdkpsil 434

RESULT 2  
AAR92315  
ID AAR92315 standard; Protein; 434 AA.  
XX  
AC AAR92315;  
XX

16-OCT-1996 (first entry)  
CORK potassium channel protein.  
CORK; potassium channel; nematode; pore-forming domain;  
transmembrane helix; N-glycosylation site; potassium-agonist;  
potassium-antagonist; drug screening; nematocide; anthelmintic;  
cardiac disorder.  
Caenorhabditis elegans.  
Key Location/Qualifiers  
Modified-site 5..7 /note= "N-glycosylation site"  
Domain 26..39 /note= "Pore-forming H5 domain"  
Misc-difference 33..35 /note= "G-Y/F-G motif"  
Modified-site 81..83 /note= "N-glycosylation site"  
Modified-site 144..146 /note= "N-glycosylation site"  
Domain 150..162 /note= "Pore-forming H5 domain"  
Misc-difference 156..158 /note= "G-Y/F-G motif"  
W09613520-A1.  
09-MAY-1996.  
25-OCT-1995; 95WO-US14364.  
31-OCT-1994; 94US-0332312.  
(AMCY ) AMERICAN CYANAMID CO.  
Pausch MH, Price LA;  
WPI; 1996-239450/24.  
N-PSDB; AAT18168.  
Potassium channel genes from Drosophila melanogaster and  
Caenorhabditis elegans - useful in assaying substances to determine  
effects on cell growth, and in inhibiting nematode and insect pests  
Claim 13; Fig 9; 79pp; English.  
This potassium channel sequence is encoded by the CORK gene from  
Caenorhabditis elegans, and has 2 pore-forming domains situated  
between hydrophobic transmembrane helix domains. The CORK sequence  
contains structural features resembling pore-forming H5 domains  
found in potassium channels. 2 Putative pore-forming H5 domains  
contain the G-Y/F-G tripeptide motif required for potassium  
selectivity. The protein contains 3 asparagine-linked glycosylation  
sites. The protein may be expressed in a heterologous host cell to  
assay substances to determine effects on cell growth. Potassium-  
agonists or potassium-antagonists identified by this method may be  
used as nematocides, anthelmintics or in therapy of cardiac  
disorders, etc.  
Sequence 434 AA;

Query Match 99.8%; Score 2290; DB 17; Length 434;  
Best Local Similarity 99.8%; Pred. No. 2.2e-227;  
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVIINRSNTYAVEQAPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDG 60  
Db 1 mviinrsntyaveqaprdkyniyvwlvilvfgvllpwnmfitiapeyyvnywfkpdg 60  
Qy 61 VETWYSKEFMGSLTIGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILILVI 120

Db 61 vetwyskefmsltigspnasinvflliaagpllyrvfapcfnvltlilivi 120  
Qy 121 VLEPTEDSNMFWFTLGMATSIINFSNGLYENSVYGVGGDFPHTYIGALLIGNNICGLLI 180  
Db 121 vleptedsmsffwvtlmgatsinfsnglyensvyvggdfphtyigallignnicglli 180  
Qy 181 TVVKIGVTVYFLNDEPKLVAIVFGISLVILVCAIALFFITKQDPYHYHHQKMEIREKA 240  
Db 181 tvvkigvfyIndepkLvaivfygislvilvcaialffitkqdfyhybhqgieireka 240  
Qy 241 ETRDPSILMTTFCNCYGFENWFCFVTLTFPPVMMVTTRGDSGFLNKINSENDEI 300  
Db 241 etdrpsilmttftncyggfnwfcfavtltfppvmvmttrgdsqflnkimsendei 300  
Qy 301 YTLTSLFVFNFAAIGSIVASKIHWPTPRYLKFAILRALFIPFFFCNVRQVTRAYPV 360  
Db 301 ytltsflvfnlfaaigsivaskihwptprylkfaillralfipfffcnvrqvtraypv 360  
Qy 361 FFESTDIFVIGGIAMSFSGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLW 420  
Db 361 ffestdifvigiamsfshgylsalamgytpnvvpshysrfaaqlsvctlmvglitgglw 420  
Qy 421 PVVIEHFVDKPSIL 434  
Db 421 pvviehfvdkpsil 434

RESULT 3  
ABB64784  
ID ABB64784 standard; Protein; 458 AA.  
XX  
AC ABB64784;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 21144.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical..  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL08887.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
PS Disclosure; SEQ ID NO 21144; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 458 AA;  
  
Query Match 28.1%; Score 645; DB 22; Length 458;  
Best Local Similarity 32.9%; Pred. NO. 9.3e-58;  
Matches 138; Conservative 83; Mismatches 178; Indels 20; Gaps 7;  
  
Qy 18 PRDKXNIYVWLIVLGVFGLLPWNMFITAPPIYNYWFKPD--GVETWYSKKEPMGSLT 74  
Db 51 pkdkfliivffllbgvgtlmpwnmfit-aksyfedfkfgpantvatevsyrthimqmg 109  
Qy 75 IGSQLPNASINNVNFIILTAGPLIYRVFAPCFNIVNLTIIILIVILEPTEDSNW--- 131  
Db 110 fasqipnlivfnwlnifvngdlttrivysiifenvillvtillami-----dssqwpqv 164  
Qy 132 FFWVTLGMATSIINFSNGLYENSVYGVGGDFPHTYIGALLIGNNICGLLITVYKIGVTVPL 191  
Db 165 ffwtmvcivllnvcngiyqntiyyivaslpikytgavvlgsgniscgftamal-icgei 223  
Qy 192 NDEPKLVAIVFGISLVILVCAIALFFITKQDPYHYHHQKMEIREKAEIRDRPSPSILW 251  
Db 224 fsskrtalsalyfvtailvllclfdtyfalpInkfrhyetlsrseksksdaqInvpyw 283  
Qy 252 TTFTNCYGFENWFCFVTLTFPPVMMVTTRGDSGFLNKINSENDEIYTLTSLFVFN 311  
Db 284 qlfkaapqlfnifltffvtlsvfpagisnvhrrspnfv---vgpdyftlvctfatfn 338  
Qy 312 LFAAIGSIVASKIHWPTPRYLKFAILRALFIPFFFCNYY--RVQTRAYPVFFESTDIFV 369  
Db 339 vfamlgslttswvqpgprflwvvpvlrlaflplfvmcnvppdsrvslavfiendwvwy 398  
Qy 370 TGGIAMSFSHGYSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLWVPIEHFV 428  
Db 399 gigiamaysgylssalgmnyabqtvtktkyttagmyaaamlitgfsylvlsgpfv 457  
  
RESULT 4  
AAy82286  
ID AAY82286 standard; Protein; 475 AA.  
XX  
AC AAY82286;  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE Rat ENT1 receptor SEQ ID NO:5.  
XX  
KW Rat; ENT1; immunosuppressant; vasotropic; thrombolytic;  
KW cytosstatic; hypotensive; antiinflammatory; analgesic; anticoagulant;  
KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;  
KW organ transplant rejection; renitis; pancreatitis; hypertension;  
KW analgesic; blood platelet coagulation inhibitor.  
XX  
OS Rattus sp.  
XX  
PN WO200012550-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 26-AUG-1999; 99WO-JP04602.  
XX  
PR 27-AUG-1998; 98JP-0241248.  
XX  
PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
PI Miyaji H, Mimura H, Kambe M, Nakagawa S;  
XX  
DR WPI; 2000-256588/22.  
XX  
DR N-PSDB; AAA08030.



Db 359 rwlpslvlarivfvpdlillcn--ikpryltwvfhdawffmaafafngylaslcmc 416  
Qy 389 YTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLPVPIEHFV 428  
Db 417 fgpkkvpaetaetagaiafflclglalgalgavfsifraiv 456

RESULT 6  
AAB15520  
ID AAB15520 standard; Protein; 456 AA.  
AC AAB15520;  
DT 14-FEB-2001 (first entry)  
DE Human ENT1 protein.  
KW Antiviral; antitumour; NBMPR-iENTP; nitrobenzylmercaptapurineriboside;  
KW insensitive, equilibrative nucleoside transporter protein; cancer;  
KW plasma membrane; facilitated diffusion; gene therapy; primer; probe.  
OS Homo sapiens.  
XX US6130065-A.  
XX 10-OCT-2000.  
XX 09-APR-1998; 98US-0058389.  
XX 11-APR-1997; 97US-0043659.  
XX (SUJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX Crawford CR, Belt JA;  
XX WPI; 2000-637839/61.  
XX New nucleic acid encoding nucleoside transporter, useful for drug  
XX screening to identify antitumour and antiviral agents and for gene  
XX therapy -  
XX Disclosure; Fig 2A-B; 55pp; English.

The invention relates to the isolation of a gene encoding a NBMPR-iENTP  
(nitrobenzylmercaptapurineriboside-insensitive, equilibrative nucleoside  
transporter protein) from HeLa cells. The protein transports nucleosides  
across the plasma membrane by a facilitated diffusion process and is  
involved in the proliferative response. The transporter is useful in  
screening to identify natural nucleoside permeants and/or their  
inhibitors or analogues, potential therapeutic agents, also for studies  
on protein structure and mechanism. Cells that express the iENTP as the  
only transporter protein are used for drug screening (especially to  
identify antitumour and antiviral nucleoside analogues), in chemotherapy  
of cancer and for selective expression of heterologous genes for gene  
therapy. The cDNA is used for recombinant expression of iENTP and as  
a source of oligonucleotides (diagnostic primers and probes, ribozymes  
and antisense sequences). Antibodies raised against iENTP are used for  
detection of the protein by usual immunoassays and as (antagonists of  
iENTP activity. This sequence represents the human ENT1 protein and  
is used for comparison with the human NBMPR-iENTP protein (AAB15517).

XX Sequence 456 AA;  
Query Match 15.08; Score 343.5; DB 21; Length 456;  
Best Local Similarity 24.38; Pred. No. 1.1e-26;  
Matches 112; Conservative 77; Mismatches 212; Indels 59; Gaps 13;  
Qy 18 PRDKYNIYWLIVLGVFGVLLPWNMTIAPYYVYVWFKPDGVE---TWYSKEFMGSLT 74  
Db 7 pddrykavwiflmgilgllpwnffmt-atqyftnrldmsqnvslvtalskdaqasaa 65

Qy 75 IGSQLP--NASINVENFLIIAGPLIYRVFA-----PVCNFIYV-----LTIILI 117  
Db 66 paaplpernsIsaifnnvmticamlplllfytlosflbhrpqsvrlgslvailvfl 125  
Qy 118 LVIVLEPTEDSMWFFWVTLMGATYSINFSNGLYENSVYGVGGDFPHYTYIGALLIGNNICG 177  
Db 126 tailvkqldalp-ffvitmikivlinsfgailgslfglagllpasytapimsqgglag 184  
Qy 178 LLITVVKIGVTVFLNDEPKLVAIVVFGISLVLLCAITALPFTTKQDFVHHQ----- 231  
Db 185 ffasvami-caiasgselsesafgyfcacaviliitiicyigrliefryyyqqklegp 243  
Qy 232 -----KGMEIREKAETDRPSPSILMTT-----FTNCYGQLFNWVFCFAVT 271  
Db 244 gegetklldiskgeepragkeesgvsvnsqptneshsikaillknslvslfscvfit 303  
Qy 272 LTIFPVMVTVTRGDSGFLNKIMSENDEIYLLTSFIVNLFPAAGISIVASKIHP--TP 329  
Db 304 igmpavtvevksslag-----sstweryfipvscfltnifdwigrsitavfmpgkds 358  
Qy 330 RYLKFAILLRALFIPFFFCNRYVOTRAY-PVFFESTDIFVIGGTAMSFSGYLSALAMG 388  
Db 359 rwlpslvlarivfvpdlillcn--ikpryltwvfhdawffmaafafngylaslcmc 416  
Qy 389 YTPNVVPSHYSRFAAQLSVCTLMVGLLTGGLPVPIEHFV 428  
Db 417 fgpkkvpaetaetagaiafflclglalgalgavfsifraiv 456

RESULT 7  
AAY82285  
ID AAY82285 standard; Protein; 475 AA.  
XX AAY82285;  
XX 19-JUN-2000 (first entry)  
DE Human ENT1 receptor SEQ ID NO:1.  
KW Human; ENT1; ENTR1; immunosuppressant; vasotropic; thrombolytic;  
KW cytostatic; hypotensive; antiinflammatory; analgesic; anticoagulant;  
KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;  
KW organ transplant rejection; renitis; pancreatitis; hypertension;  
KW analgesic; blood platelet coagulation inhibitor.  
XX Homo sapiens.  
XX WO200012550-A1.  
XX 09-MAR-2000.  
XX 26-AUG-1999; 99WO-JP04602.  
XX 27-AUG-1998; 98JP-0241248.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Miyaaji H, Mimura H, Kambe M, Nakagawa S;  
XX WPI; 2000-256588/22.  
XX N-FSDB; AAA08027.  
XX Novel nucleoside transporter polypeptide, useful in drug development  
XX for agents in treating e.g. ischemia, cerebral embolism, rejection in  
XX organ transplant, malignant tumors, renitis, pancreatitis and  
XX hypertension -  
XX Claim 1; Page 70-74; 94pp; Japanese.  
XX The present invention describes human and rat ENT1 receptors (ENTR1),  
XX with some amino-acids deleted, substituted or added, but which retain  
XX nucleoside transport activity, and can be used in drug development for  
XX agents in treating e.g. ischaemia, cerebral embolism, rejection in organ



XX DE Equilibrative nucleoside transport protein (iENTP).

XX KW Equilibrative nucleoside transport protein; iENTP; NBMPR; transporter;

XX KW nitrobenzylmercaptapurine riboside; antiviral; antitumour; screening;

XX KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;

XX KW adenosine deaminase; factor VIII.

XX OS Homo sapiens.

XX PH Location/Qualifiers

FT Domain 13..27

FT /note= "transmembrane domain TM1"

FT Domain 71..93

FT /note= "transmembrane domain TM2"

FT Domain 100..117

FT /note= "transmembrane domain TM3"

FT Domain 125..144

FT /note= "transmembrane domain TM4"

FT Domain 162..185

FT /note= "transmembrane domain TM5"

FT Domain 194..214

FT /note= "transmembrane domain TM6"

FT Domain 293..312

FT /note= "transmembrane domain TM7"

FT Domain 326..341

FT /note= "transmembrane domain TM8"

FT Domain 361..379

FT /note= "transmembrane domain TM9"

FT Domain 393..413

FT /note= "transmembrane domain TM10"

FT Domain 432..452

FT /note= "transmembrane domain TM11"

XX WO9846749-A1.

XX PD 22-OCT-1998.

XX PF 10-APR-1998; 98WO-US07283.

XX PR 09-APR-1998; 98US-0058389.

XX PR 11-APR-1997; 97US-0838845.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Belt JA, Crawford CR, Patel DH;

XX DR WPI; 1998-594576/50.

XX DR N-PSDB; AAV69698.

XX PS New equilibrative nucleoside transport protein insensitive to

XX PT nitrobenzylthio-purine riboside - useful for, e.g. identifying

XX PT specific therapeutic nucleoside analogues and in gene therapy to

XX PT protect transduced cells against ablative chemotherapy

XX Claim 13; Pages 71-72; 114pp; English.

XX This represents an equilibrative nucleoside transport protein (iENTP),

XX which is insensitive to nitrobenzylmercaptapurine riboside (NBMPR). Cells

XX transformed with a construct containing the iENTP nucleic acid can be

XX used to produce the protein recombinantly. iENTP is used to identify

XX specific ligands (particularly antiviral and antitumour nucleoside

XX analogues that are preferentially transported into cells) and to raise

XX antibodies. Cells in which iENTP provides all available transport

XX activity are used: (a) to identify permeants of iENTP and (b) to screen

XX specific inhibitors of iENTP (potential drugs). Fragments of the iENTP

XX nucleic acid are used, as probes, primers, antisense molecules, and

XX ribozymes for therapy or diagnosis, and knockout mice in which both

XX alleles encoding iENTP contain an inactivating defect are also useful for

XX drug screening. Cells that have been transduced with iENTP nucleic acid

XX ex vivo are used particularly, for cancer chemotherapy. Vectors in which

XX the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine

XX deaminase and factor VIII) can be used for gene therapy.

XX SQ Sequence 456 AA;

Query Match 14.2%; Score 326.5; DB 19; Length 456;

Best Local Similarity 22.9%; Pred. No. 6.3e-25;

Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKYNIVVWLVILVGVGLLPNNMFITIAPEY-----VNWFKPDGYE 62

DB 7 prdsynlvigisffilgltclpwnffitaipfqrllqagaginstarilstnhtgpedaf- 65

QY 63 TWYSKEFMGSLTIGSOLPNASINVFNLFIITAGPLIYRVFAPVCFNIVNLTIILIVL 122

DB 66 -----nfnwnvtllsqpllllftllnsflyqcvpetvrlg----sllailllfaltaal 116

QY 123 EPTEDSMSWFFWVTLGMATSNFNSGLYENSIVGVGGDPHPTIYIGALLIGNICGLLITV 182

DB 117 vkvdmspgpffsittmasvcfinsfsaviqgsifgqlgtmpstystlffsggqlagifaa 176

QY 183 VKI-----GVTYFLNDEPKLVAIVYF---GISLVILLVCAIALFFITKODFY---HYHHQ 231

DB 177 amlslmasgv-----daetsalgylfitpcvgilmsivcylslphlkkfaryylankssa 230

QY 232 KGMEIREKA-----ETDRPSPSILMTTFTNC 257

DB 231 qaeleltaellqsdengipsspqkvaltdldlekepesepepqpqkpsvftvfqki 290

QY 258 YQOLFNVWFCEFAVTLTIPTVMVTTRGDSGFLNKIMSENDEIYTLTSLFVFNLFAAIG 317

DB 291 witaiclvlvftvtlsvfipaitamvtslts-----pgkwsqffnfpiccllfnmdwlg 344

QY 318 SIVASKIHWP--TPRYLKFAIILRALFIPFFFCNVTRVOTRAYPVFFESTDIFVIGGIAM 375

DB 345 rsitsyflwpdedsrllpllvclrfilvplfmlchvpqrsr-lpilfpqdayfitmllf 403

QY 376 SFSHGYSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTG 417

DB 404 avsnngylvsiltnciaprqvlppherevagalmftflalglscg 445

RESULT 10

AAW69558

ID AAW69558 standard; Protein; 456 AA.

XX AC AAW69558;

XX DT 13-OCT-1998 (first entry)

XX DE Human equilibrative nucleoside transporter 2.

XX KW Human; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;

XX KW rENT2; coronary; cerebrovascular anoxia; viral infection; cancer.

XX OS Homo sapiens.

XX PN WO9829437-A2.

XX PD 09-JUL-1998.

XX PF 30-DEC-1997; 97WO-IB01657.

XX PR 03-NOV-1997; 97US-0064004.

XX PR 30-DEC-1996; 96US-0034083.

XX PA (UYAL-) UNIV ALBERTA.

XX PA (UYLE-) UNIV LEEDS.

XX PI Baldwin SA, Cass CE, Young JD;

XX DR WPI; 1998-388035/33.

XX DR N-PSDB; AAV40279.

PT Newly isolated equilibrative nucleoside transporter protein(s) and  
 PT gene(s) - used to develop products for treating disorder(s)  
 PT associated with the transporter(s) and for use with nucleoside  
 PT drug(s)

XX Claim 5; Fig 20; 97pp; English.

XX The present sequence represents a substantially purified equilibrative  
 CC nucleoside transporter (ENT), human ENT2 (iENT2). ENTs can transport a  
 CC variety of purines and pyrimidines, including adenosine, uridine,  
 CC guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are  
 CC bidirectional, they transport a suitable permeant both into and out of  
 CC cells. ENTs can be used as a tool for the development of new nucleoside  
 CC drugs. Products from the present invention can be used for treating a  
 CC subject having a disorder associated with an ENT. They can also be used  
 CC with nucleoside drugs in the treatment of e.g. coronary or  
 CC cerebrovascular anoxia, viral infection or cancer. The products (e.g.  
 CC antibodies and oligonucleotides) hybridising to nucleic acid sequences  
 CC encoding ENTs) can also be used for detection and drug screening.

XX Sequence 456 AA;

Query Match 14.2%; Score 326.5; DB 19; Length 456;  
 Best Local Similarity 22.9%; Pred. No. 6.3e-25;  
 Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

Qy 18 PRDKYNIYVWLVILVFGVLLPWNFFITIAPEY-----VNWFKPDGVE 62  
 Db 7 prdsyhlvgifilglgtllpwnffitaipfqrarlagagnstarilsthtgpedaf- 65  
 Qy 63 TWYSKEFGSLTIGSOLPNASINVENLFIITAGPLIYRVFAPVCNIVNLITLILVIVL 122  
 Db 66 -----nfnnwvllsqplllftlinsflyqcvpetvilg-----sllaillfaltaal 116  
 Qy 123 EPTEDSMWFFWVTIGMATSNFNSGLYENSIVYGGDFPHYIGALLIGNNICLLITV 182  
 Db 117 vkvdmspgpfisitmasvcfinsfsavlgqslfglqgmpstystflisggglagifaa 176  
 Qy 183 VKI-----GVTYFLNDEPKLVAIVF---GISLVLLVCAITAFITTKQDFY---HYHHQ 231  
 Db 177 amllsmasgv-----daetsalgyfltpcygilmisvcyislphikfaryylankssqa 230  
 Qy 232 KGMETREKA-----ETDRSPSILWTTFTNC 257  
 Db 231 qaqeletkaellqsdengipspqkvaltdldlekepesepepdkpgkpsvftvfaki 290  
 Qy 258 YQGFENWFCFAVLTITPPVMVTTTRGDSGFLNKIMSENDEIYTLTSLFVFNLFAAIG 317  
 Db 291 wltalcvlvftvtlsvfpaitamvts-----pgkwsqffnppiccfllfnmdwlg 344  
 Qy 318 SIVASKIHWP--TPRYLKFAILLRALTFPEFFFCNRYQTRAYPVFFESTDIEVIGGAM 375  
 Db 345 rsitsyflwddesrllpllcrlfrifvplfmlchvqrsr-lpllfpqdayfifcmlf 403  
 Qy 376 SFSHGYSALAMGYTPNVPSHYSRFAQLSVCTLMVGLLTG 417  
 Db 404 avsnqylvslmclaprvlpherevagalmftflalglscg 445

RESULT 11

ID AAB15517  
 AC AAB15517 standard; Protein; 456 AA.

XX AAB15517;

XX 14-FEB-2001 (first entry)

XX Human NBMPR-iENTP protein.

DE Antiviral; anticancer; NBMPR-iENTP; nitrobenzylmercaptopurineriboside;  
 KW insensitive, equilibrative nucleoside transporter protein; cancer;  
 KW plasma membrane; facilitated diffusion; gene therapy; primer; probe.

XX Homo sapiens.  
 OS US6130065-A.  
 PN 10-OCT-2000.

XX 09-APR-1998; 98US-0058389.  
 XX 11-APR-1997; 97US-0043659.  
 XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 XX Crawford CR, Belt JA;

XX WPI; 2000-637839/61.  
 XX N-PSDB; AAA95722.

XX New nucleic acid encoding nucleoside transporter, useful for drug  
 PT screening to identify antitumour and antiviral agents and for gene  
 PT therapy -

XX Claim 1; Fig 2A-B; 55pp; English.

XX This sequence represents nitrobenzylmercaptopurineriboside-insensitive,  
 CC equilibrative nucleoside transporter protein (NBMPR-iENTP) isolated from  
 CC HeLa cells. The protein transports nucleosides across the plasma membrane  
 CC by a facilitated diffusion process and is involved in the proliferative  
 CC response. The transporter is useful in screening to identify natural  
 CC nucleoside permeants and/or their inhibitors or analogues, potential  
 CC therapeutic agents, also for studies on protein structure and mechanism.  
 CC Cells that express the iENTP as the only transporter protein are used  
 CC for drug screening (especially to identify antitumour and antiviral  
 CC nucleoside analogues), in chemotherapy of cancer and for selective  
 CC expression of heterologous genes for gene therapy. The cDNA is used  
 CC for recombinant expression of iENTP and as a source of oligonucleotides  
 CC (diagnostic primers and probes, ribozymes and antisense sequences).  
 CC Antibodies raised against iENTP are used for detection of the protein  
 CC by usual immunoassays and as (ant)agonists of iENTP activity.

XX Sequence 456 AA;

Query Match 14.2%; Score 326.5; DB 21; Length 456;  
 Best Local Similarity 22.9%; Pred. No. 6.3e-25;  
 Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

Qy 18 PRDKYNIYVWLVILVFGVLLPWNFFITIAPEY-----VNWFKPDGVE 62  
 Db 7 prdsyhlvgifilglgtllpwnffitaipfqrarlagagnstarilsthtgpedaf- 65  
 Qy 63 TWYSKEFGSLTIGSOLPNASINVENLFIITAGPLIYRVFAPVCNIVNLITLILVIVL 122  
 Db 66 -----nfnnwvllsqplllftlinsflyqcvpetvilg-----sllaillfaltaal 116  
 Qy 123 EPTEDSMWFFWVTIGMATSNFNSGLYENSIVYGGDFPHYIGALLIGNNICLLITV 182  
 Db 117 vkvdmspgpfisitmasvcfinsfsavlgqslfglqgmpstystflisggglagifaa 176  
 Qy 183 VKI-----GVTYFLNDEPKLVAIVF---GISLVLLVCAITAFITTKQDFY---HYHHQ 231  
 Db 177 amllsmasgv-----daetsalgyfltpcygilmisvcyislphikfaryylankssqa 230  
 Qy 232 KGMETREKA-----ETDRSPSILWTTFTNC 257  
 Db 231 qaqeletkaellqsdengipspqkvaltdldlekepesepepdkpgkpsvftvfaki 290  
 Qy 258 YQGFENWFCFAVLTITPPVMVTTTRGDSGFLNKIMSENDEIYTLTSLFVFNLFAAIG 317  
 Db 291 wltalcvlvftvtlsvfpaitamvts-----pgkwsqffnppiccfllfnmdwlg 344  
 Qy 318 SIVASKIHWP--TPRYLKFAILLRALTFPEFFFCNRYQTRAYPVFFESTDIEVIGGAM 375













**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 17:35:27 ; Search time 3087.27 Seconds  
(without alignments)  
6068.071 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggtaataatcaccggtac.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Using first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	96.2	6.9	551	10	Bj119282	Bj119282 Bj119282
C 2	96.2	6.9	669	10	Bj147020	Bj147020 Bj147020
C 3	86.6	6.2	533	9	AU222343	AU222343 AU222343
4	77.6	5.6	332	9	AU201199	AU201199 AU201199
C 5	75.6	5.4	710	9	AU216992	AU216992 AU216992
6	75.2	5.4	648	9	AW453425	AW453425 SWOV3MCA
7	73.8	5.3	327	9	AU109268	AU109268 AU109268
8	73.8	5.3	642	9	AU082929	AU082929 SWAMCAG12
9	70.8	5.1	495	9	AU209436	AU209436 AU209436
10	70.2	5.1	360	10	CI3829	CI3829 CI3829
C 11	67.8	4.9	300	9	AU113471	AU113471 AU113471
C 12	57.6	4.1	300	9	AU115402	AU115402 AU115402
C 13	57.4	4.1	300	9	AU114995	AU114995 AU114995
14	56.2	4.0	997	12	CNS005TE	AL060767 Drosophila
15	54.4	3.9	258	9	AA406898	AA406898 MBACFC27F0
C 16	54.4	3.9	543	10	BI501984	BI501984 rm07f05.Y
17	53	3.8	656	9	AW409482	AW409482 SWOV3MCA

18	50.4	3.6	345	9	AU109592	AU109592 AU109592
19	49	3.5	606	9	AA948920	AA948920 LD27618.5
C 20	48.8	3.5	901	12	CNS0760F	AL430789 clone XBA
21	48.2	3.5	1201	12	CNS0010J	AL054622 Drosophila
22	47.6	3.4	274	9	AA842090	AA842090 MBACFC380
23	47	3.4	228	9	AI784877	AI784877 SWAMCAG40
24	46.8	3.4	729	9	AI239052	AI239052 GH15222.5
25	46.2	3.3	360	10	R03450	R03450 pk06g03.r1
26	46.2	3.3	373	10	C43216	C43216 C43216 Yuj1
27	45.8	3.3	374	9	BB181657	BB181657 BB181657
28	45.6	3.3	350	9	AU111089	AU111089 AU111089
29	45.6	3.3	351	9	AU111161	AU111161 AU111161
30	45.6	3.3	738	12	AQ842781	AQ842781 CpG1299A
C 31	45.6	3.3	759	12	AQ254298	AQ254298 CpG0751B
32	45	3.2	375	10	C44886	C44886 C44886 Yuj1
33	44.6	3.2	340	9	AU111560	AU111560 AU111560
34	44.6	3.2	340	9	AU111734	AU111734 AU111734
C 35	44.6	3.2	510	12	AQ254261	AQ254261 CpG0732A
C 36	44.4	3.2	429	9	AW829621	AW829621 ra42d09.y
C 37	44.4	3.2	448	9	AW829541	AW829541 ra41q09.y
C 38	44.2	3.2	333	9	AI431074	AI431074 mj58g09.x
C 39	44.2	3.2	441	10	BF459052	BF459052 UI-M-B21-
40	44.2	3.2	442	9	BB750597	BB750597 BB750597
41	44.2	3.2	459	9	AI595433	AI595433 mj58g09.y
42	44.2	3.2	475	9	AA058117	AA058117 mj58g09.r
C 43	44.2	3.2	566	9	AW333515	AW333515 S22G3 AGS
C 44	44.2	3.2	626	9	AW048063	AW048063 UI-M-BH1-
45	44.2	3.2	649	9	BB169820	BB169820 BB169820

ALIGNMENTS

RESULT 1  
Bj119282  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
BASE COUNT  
ORIGIN  
Query Match

551 bp mRNA linear EST 23-JAN-2002  
unpublished oligo-capped cDNA library, C. elegans L1 stage  
Caenorhabditis elegans cDNA clone yk1241g07 5', mRNA sequence.  
Bj119282  
Caenorhabditis elegans  
EST.  
Bj119282.1 GI:18279408  
Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 551)  
Kohara, F., Shin-I, I., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.  
and Sugano, S.  
A complementary view of the C.elegans genome  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
I. .551  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk1241g07"  
/clone\_lib="unpublished oligo-capped cDNA library, C.  
elegans L1 stage"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="L1"

117 a 152 c 94 g 187 t 1 others  
6.9%; Score 96.2; DB 10; Length 551;

Best Local Similarity 53.6%; Pred. No. 1.8e-09;  
Matches 200; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy 911 tcacagttctcgtctcaattgttcgctgcgagtgatccatagttgcttccaaga 970
Db 27 TGACCACATTCCTTCAATTCAACGTTGTCGCTTTTATTGGATCAATTTGCGCGGAAGAA 86
Qy 971 ttcactggcgacaccccgcttaccctcaaatgtgccataatcttgctgctctttccatc 1030
Db 87 AGCAATGGCTCGCGGACACAGCTCGGATTCAGCTTCCAGTCTACCTCCGTCCTCTATATTC 146
Qy 1031 catttcttcttcgcaactatcgtgtccagacgctgcttatcctgtttcttttgagt 1090
Db 147 CATTTCTTCATCTCTCTGCAACTATCTCCCTGAGACCGGTTCACTTCCCGTCTTTTTCGAAT 206
Qy 1091 ctactgacatctttgtgattggtgaattgccatgctctttttcacatgatacctcagcg 1150
Db 207 CCACCTGGCTTTTCGTCATATTCGCGGCTCGATGAGCTTTGGAGTGATATTCCTCGG 266
Qy 1151 ctctggcaatgggatacacctccaaacgtgctgccatctcactactcaagattgcccgtc 1210
Db 267 GACTCGCATGATGATACACCTCGAAGACTGTTGATCCATCGAAGGCTCAGGTCGCTGGAA 326
Qy 1211 agcttccggttgcaactcttattgttgcccttcacacggtggtggtggtggttggta 1270
Db 327 TGATGCGCGGATTCCTTCATCTCTGGAATTGCTCTGCTGCTGATCTTCACAATGGTCA 386
Qy 1271 ttgagcaactcgt 1283
Db 387 TCAAGATGTCGT 399
```

## RESULT 2

LOCUS BJ147020/c BJ147020 669 bp mRNA linear EST 24-JAN-2002  
DEFINITION BJ147020 unpublished oligo-capped cDNA library, C. elegans L1 stage  
Caenorhabditis elegans cDNA clone yk1241g07 3', mRNA sequence.  
ACCESSION BJ147020  
VERSION BJ147020.1 GI:18315005  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.  
1 (bases 1 to 669)  
AUTHORS and Sugano, S.  
TITLE A complementary view of the C.elegans genome  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

## FEATURES

source  
1. 669  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk1241g07"  
/clone\_lib="unpublished oligo-capped cDNA library, C.  
elegans L1 stage"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="L1"  
BASE COUNT 240 a 106 c 170 g 151 t 2 others  
ORIGIN

Query Match

Best Local Similarity 53.6%; Pred. NO. 1.7e-09;  
Matches 200; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

Matches 200; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy 911 tcacagttctcgtctcaattgttcgctgcgagtgatccatagttgcttccaaga 970
Db 659 TGACCACATTCCTTCAATTCAACGTTGTCGCTTTTATTGGATCAATTTGCGCGGAAGAA 600
Qy 971 ttcactggcgacaccccgcttaccctcaaatgtgccataatcttgctgctctttccatc 1030
Db 599 AGCAATGGCTCGCGGACACAGCTCGGATTCAGCTTCCAGTCTACCTCCGTCCTCTATATTC 540
Qy 1031 catttcttcttcgcaactatcgtgtccagacgctgcttatcctgtttcttttgagt 1090
Db 539 CATTTCTTCATCTCTCTGCAACTATCTCCCTGAGACCGGTTCACTTCCCGTCTTTTTCGAAT 480
Qy 1091 ctactgacatctttgtgattggtgaattgccatgctctttttcacatgatacctcagcg 1150
Db 479 CCACCTGGCTTTTCGTCATATTCGCGGCTCGATGAGCTTTGGAGTGATATTCCTCGG 420
Qy 1151 ctctggcaatgggatacacctccaaacgtgctgccatctcactactcaagattgcccgtc 1210
Db 419 GACTCGCATGATGATACACCTCGAAGACTGTTGATCCATCGAAGGCTCAGGTCGCTGGAA 360
Qy 1211 agcttccggttgcaactcttattgttgcccttcacacggtggtggtggtggttggta 1270
Db 359 TGATGCGCGGATTCCTTCATCTCTGGAATTGCTCTGCTGCTGATCTTCACAATGGTCA 300
Qy 1271 ttgagcaactcgt 1283
Db 299 TCAAGATGTCGT 287
```

## RESULT 3

LOCUS AU222343/c AU222343 533 bp mRNA linear EST 17-JUL-2001  
DEFINITION AU222343 unpublished oligo-capped cDNA library, stage L1  
Caenorhabditis elegans cDNA clone yk1015b04 3', mRNA sequence.  
ACCESSION AU222343  
VERSION AU222343.1 GI:14860500  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.  
1 (bases 1 to 533)  
AUTHORS and Sugano, S.  
TITLE A complementary view of the C.elegans genome  
JOURNAL Unpublished (2001)  
COMMENT Contact: Yuji Kohara  
Genome Biology Lab.  
National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.

## FEATURES

source  
1. 533  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk1015b04"  
/clone\_lib="unpublished oligo-capped cDNA library, stage  
L1"  
/sex="Hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="L1"  
BASE COUNT 191 a 101 c 100 g 141 t  
ORIGIN

Query Match

Best Local Similarity 49.3%; Pred. NO. 1.5e-07;  
Matches 255; Conservative 0; Mismatches 259; Indels 3; Gaps 1;



```
Qy 808 gttactctcaaatcttccctgttatgatgaccgtttaccactcgtggagattccggcttc 867
Db 514 GTTCTCTCTCATCTTTCTCTGCAATGACTGATGATGTGTACTCTGATCTTGAACCTTACAAAT 455
Qy 868 ctaacaaaaatagtctgaaacagatgaaatctacacatttgctcacaagtttctctcgtc 927
Db 454 GGAATAAATCTGTTGTTTGGAGATAGTTTATTTCTCCG---GATAAATCTACTTCTCTCAAT 398
Qy 928 ttcaattgttcgtcgatggatccatagttgcttcccaagattcactggccgacaccc 987
Db 397 TTCAATTTATTGATGATGGATGGATCATCTCTTGCCAAATATGTTCAAATCCCATCAGAG 338
Qy 988 cgttaccctcaaatggcataatcttgctgctctcttccattccattctcttctctcgc 1047
Db 337 AAATATTATTGATGCGGAGTTCCTTCGAACCGTATTCATTCATTCATTTTCTCTGT 278
Qy 1048 aactatcgttcacagacgctgctctatcctctgttcttcttgtagtctactgacattttgtg 1107
Db 277 AATTATCTCCAAATACTAGAAGATGGCTGTGTGGTTTAAAGAAATGAATGGTGTCACT 218
Qy 1108 attggtgaattgccatctcttttccatcgtatcactcagcgtctgcaatgggatac 1167
Db 217 ATTGATCTACCAATATGGCAATTTACTTGTGTTATATGACGAGTTTAGCGCTGATTTAT 158
Qy 1168 actcaaacgctgctgccatctcactcactcaagatttgcgctcagcttctcgttttgcact 1227
Db 157 ACACCAAGCAAGTCCAGCTAGATATCAGAAGCTAAGTGGAAATGCTTGCATCAATTTTC 98
Qy 1228 cttatggttgcctctcaccggtggtcgtggtgcccgttgttattgagcacttcgtgac 1287
Db 97 CTAATGCTCGGAATTTCTATCGGAGTCCAGTACACCATTCGTGTCATGGCGGTGGAC 38
Qy 1288 aagccaagtatctataaattattatacattagatt 1324
Db 37 TCGATAGGAAGTAGAAGACATGACTGTATAAATTT 1

RESULT 4
LOCUS AU201199 332 bp mRNA linear EST 17-JUL-2001
DEFINITION AU201199 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk789e10 5', mRNA sequence.
ACCESSION AU201199
VERSION AU201199.1 GI:14829871
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 332)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. 332
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk789e10"
/clone_lib="unpublished oligo-capped cDNA library, stage
L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"

FEATURES
source
1. 332
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk789e10"
/clone_lib="unpublished oligo-capped cDNA library, stage
L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"

BASE COUNT 251 a 103 c 181 g 172 t 3 others
ORIGIN

Query Match 5.4%; Score 75.6; DB 9; Length 710;
Best Local Similarity 51.6%; Pred. No. 2.4e-05;
Matches 199; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

Qy 949 ggatccatagttgcttcccaagattcactggccgacacccggttaacctaaatttgcata 1008
Db 709 GGATCAATCGTCGCGGGAGAAAGCAATGTCGCCGCCGCGGAAACAAAGCTCTGGATCCAGTC 650
Qy 1009 atcttgctgctcttcttccattccattcttcttctctcgaactatcgttccagacggt 1068
Db 649 TACCTCCGTCCTCTACATCCCATCTTCATCTTCTGCAACTACTCCCTGAGACCCGC 590
Qy 1069 gcttatcctgtttcttcttgagctactgacatttttggatgggggaattgcttct 1128
```

```

Db 589 ACACGCCGGTCTTCTGCAATCCACGTGGCTTTTCATCATTTGTGGCGCATCGATGAGC 530
Qy 1129 ttttcacatgatacctcagcgtctggaatggatagatacactccaaacgtcgtgcattc 1188
Db 529 TTCGGAAGTGGATACATTCAGGACTCCCATGATGTACACTTCAAGACTGTGATCCA 470
Qy 1189 cactactcaagattggcgcgtcagctttccgcttgacactcttatgttgcccttcaccc 1248
Db 469 TCAAGGCTCAGGTGGCTGGAATGATGGCTGGATTTTCCCTCATTTTCAGGAATCGTCTCA 410
Qy 1249 ggtgacctgtggccgttggttattgagcaacttcgtggacaagcccaagtattctataata 1308
Db 409 GGTCTCATTTTTCACATGATGCTATCAAG--TTTGTCGTCACTGCCTCAAAATTTTCAAAA 353
Qy 1309 ttattagcattagatatacttgta 1334
Db 352 CTCTTAACATTAAGACACCTCATTA 327

RESULT 6
AW453425 648 bp mRNA linear EST 17-FEB-2000
LOCUS SWOV3MCAM41C06SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SU96MLW-Ovml3) Onchocerca volvulus cDNA clone SWOV3MCAM41C06 5',
mRNA sequence.
ACCESSION AW453425
VERSION AW453425.1 GI:6994211
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
1 (bases 1 to 648)
Williams S.A., Lizotte-Waniewski M., Laney S. and Lustigman S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
Location/Qualifiers
1..648
/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SWOV3MCAM41C06"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SU96MLW-Ovml3)"
/dev_stage="molting L3"
/lab_host="X1-Blue MRF"
/notes="Vector: Lambda Uni-ZAP XR; Site1: Eco RI; Site2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3). 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by Rnase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10E6 independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustigmen@bc.org)."

```

```

BASE COUNT 168 a 126 c 120 g 234 t
ORIGIN
Query Match 5.4%; Score 75.2; DB 9; Length 648;
Best Local Similarity 48.9%; Pred. NO. 2.9e-05;
Matches 305; Conservative 0; Mismatches 298; Indels 21; Gaps 3;
Qy 30 tgccttgagcaggaagcatttccaagagacaagtaacaatttctactggtcgcat 89
Db 28 TGATATGGTGTTCGACAAACCTCCCAAGACAAGTACAACTGTGTTTATTCATTTACT 87
Qy 90 tcttcttggtgattcggagttcttctgccaatggaatattgttctactatcgccctgagta 149
Db 88 ACTTCACGGGATAGCGCTTTCATCCCGTGGAACTGTTTCATCACAATTCCTCCATCTTA 147
Qy 150 ttatgtgaattattggttc---aaacggatggcctggagacatgg-----ta 194
Db 148 CTACGTGAACATATAAATTCGTGGAATGAATGCCGATGATAGCGTGCATAAAAGTGATTA 207
Qy 195 itcgaagaattcatgggaatctttgacgattggctcacaacttccaaacgaagcattaa 254
Db 208 TCGCGTTCATTTTGTCTATATCTTGGTCTTGTCTTCACAGATACCGAATTTGCTACTCAA 267
Qy 255 tgtttcaacctgttccctcaattattgtggtggtccctgatctaccgcgtcttctgcccgt 314
Db 268 TCTGATCAACTTATTCGTCGACATCAAAAGGGGATCTCAGACGACGAATTAGCTTTTCGCT 327
Qy 315 ttgcttcaacatcgtaacacctgacaatcattctctctcgtcattgttctggaagcccaac 374
Db 328 TTTTGTGTTGGCGGTATCACTCTAATAACTCTCATATATCACTCTTATTAAT------AC 381
Qy 375 tgaagattccatctcctggttttctggttaactcttggaaatggcgacticaaat 434
Db 382 TTCATCATATGATCGGTGATTTTCTTTATTAACAAGACCAAGGTTGTTATTAATGC 441
Qy 435 tagcaatgggctatatgaaaactcggtttatggagttggtggcgatatttccgcacaccta 494
Db 442 TGCTAATGGTATTATCAAAAGTTCACTTTATGTTTAACTGCCACCTTCCCGCCCTCAGTA 501
Qy 495 cattggcgtctcttgattggaacaacatttgcgattgctgataacaggttgtaaaat 554
Db 502 TACCAATGCATTGATGACTGGGCAACAACATATATGCGGGACATTTGTGTCGCGTGAATA 561
Qy 555 cggagtgacctatttctgaaatgatgacgtcaaacatttctgcaatcgtctatttcggcat 614
Db 562 TCATAACACTTTGTTGGTGCAAAAATATATGATGGATGGCGGCTTCTCTATTTTGTGAT 621
Qy 615 atcgttggtgatcctctcgtggtg 638
Db 622 ATCTTGTGTAACAAGTTTGTGTTG 645

RESULT 7
AW109268 327 bp mRNA linear EST 19-OCT-2000
LOCUS AW109268
DEFINITION AW109268 unpublished oligo-capped cDNA library Caenorhabditis
elegans cDNA clone yk7084 5', mRNA sequence.
ACCESSION AW109268
VERSION AW109268.1 GI:10922835
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 327)
Thierry-Mieg J., Thierry-Mieg D., Suzuki Y.,
Kohara Y., Shin I.T.,
and Sugano S.
REFERENCE A complementary view of the C. elegans genome
AUTHORS Unpublished (2000)
JOURNAL Contact: Yuji Kohara
COMMENT Genome Biology Lab.
National Institute of Genetics

```





```
BASE COUNT      114 a      50 c      59 g      77 t
ORIGIN

Query Match      4.9%; Score 67.8; DB 9; Length 300;
Best Local Similarity 55.8%; Pred. No. 0.001;
Matches 129; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 1024 ttcattcattcttcttctgcaactatcggtccagacgctgtcttatccctgttttc 1083
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 TTCATTCATCTCTCATGTTTCAGCAACTATTACCAACATTCCTGCTACTATGGGAGTATTA 240

Qy 1084 tttagtctactgacatttttgatggttggaattgccatgtctttttcacatggatac 1143
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 TTCACCAACGAATGGATCTCTCTTTTGGTAACACATATTATGGCATTACACAAGTGGATAT 180

Qy 1144 ctcaagcctctggcaatgggatacacactccaaacgctgcccactactcaagaattt 1203
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 TCTCATCTTTGGGAATGATGATATACACCACCGTCTGCCCCACCGGAATACTCAAAATTA 120

Qy 1204 gcoqctcagctttcgtttgcactcttatggttgcccttctcaccggtggc 1254
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 GCTGCTCAAGTCTGCATGCTCTCTGCTTCTGCTGTTGCTATCTACTGCTGTGTC 69

RESULT 12
AUI15402/c
LOCUS
DEFINITION
AUI15402 unpublished oligo-capped cDNA library Caenorhabditis
ACCESSION
AUI15402
VERSION
AUI15402.1 GI:10928969
KEYWORDS
EST.
SOURCE
Caenorhabditis elegans.
ORGANISM
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
1 (bases 1 to 300)
AUTHORS
Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE
A complementary view of the C. elegans genome
JOURNAL
Unpublished (2000)
COMMENT
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
source
1..300
Location/Qualifiers
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk732c11"
/clone_lib="unpublished oligo-capped cDNA library"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"

BASE COUNT      93 a      58 c      56 g      83 t
ORIGIN

Query Match      4.1%; Score 57.6; DB 9; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.12;
Matches 126; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1015 cgtgctcttttcattccattcttcttctgcaactatcggtcccagacgctgcttat 1074
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CGAACCGTATTCATTCATTTATCTTTCTGTGTAATATCGTCCAGATACATAGAGATGG 240

Qy 1075 cctgtttcttgagtctactgacatttttgatggttggaattgcatgtcttttca 1134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



ORIGIN

Query Match 3.9%; Score 54.4; DB 9; Length 258;  
Best Local Similarity 61.1%; Pred. No. 0.53;  
Matches 88; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy	42	ggaagcatttccaagagagacaagtacaatatgtctactggctcgtctcattcttgttgatt	101
Db	19	GGATAAACCTCCTAAAGACAAATATAATGCTGTCTATTTTCACTTTTACTACTTCACGGGAT	78
Qy	102	cggagttcttctgcatggaatatgttcattactatcgccctgagtgattatgtgaatta	161
Db	79	AGGTGTTTGTGATGCCATGGACATGTTTCATCACAAATTCCTCCCTCTTACTACGTGGATTA	138
Qy	162	ttggttcaaacccggatgagcgtgga	185
Db	139	TAAATTTGTGGAGGTGAGCGCGGA	162

Search completed: July 23, 2002, 17:35:40  
Job time: 3563 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 19:15:32 ; Search time 4778.58 Seconds  
(without alignments)  
6078.382 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggtaataatcaacgcatc.....ttattataaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:  
1: gb\_ba:  
2: gb\_mt:  
3: gb\_in:  
4: gb\_om:  
5: gb\_ov:  
6: gb\_pat:  
7: gb\_ph:  
8: gb\_pi:  
9: gb\_pr:  
10: gb\_ro:  
11: gb\_sts:  
12: gb\_sy:  
13: gb\_un:  
14: gb\_vi:  
15: em\_ba:  
16: em\_fun:  
17: em\_hum:  
18: em\_in:  
19: em\_inu:  
20: em\_om:  
21: em\_or:  
22: em\_ov:  
23: em\_pat:  
24: em\_ph:  
25: em\_pi:  
26: em\_ro:  
27: em\_sts:  
28: em\_un:  
29: em\_vi:  
30: em\_htg\_hum:  
31: em\_htg\_inv:  
32: em\_htg\_other:  
33: em\_htg\_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

RESULT 1  
AX224554  
LOCUS AX224554  
DEFINITION Sequence 36 from Patent WO0161006.  
ACCESSION AX224554  
VERSION AX224554.1 GI:15554749  
KEYWORDS  
SOURCE  
ORGANISM  
Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 1394)  
AUTHORS Pausch, M. H.  
TITLE Two pore potassium channels, ~~Caenorhabditis elegans~~  
and methods of using same  
JOURNAL Patent: WO 0161006-A 36 23-AUG-2001;  
Basf Corporation (US)  
FEATURES  
Location/Qualifiers  
Source 1..1394  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
BASE COUNT 347 a 317 c 270 g 460 t  
ORIGIN

ALIGNMENTS

RESULT	1	AX224554	1394 bp	DNA	linear	PAT 10-SEP-2001
LOCUS	AX224554	Sequence 36 from Patent WO0161006.				
DEFINITION	AX224554	Sequence 36 from Patent WO0161006.				
ACCESSION	AX224554	Sequence 36 from Patent WO0161006.				
VERSION	AX224554.1	GI:15554749				
KEYWORDS						
SOURCE						
ORGANISM						
Caenorhabditis elegans.						
Caenorhabditis elegans						
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;						
Rhabditoides; Rhabditidae; Peloderinae; Caenorhabditis.						
REFERENCE 1 (bases 1 to 1394)						
AUTHORS Pausch, M. H.						
TITLE Two pore potassium channels, <del>Caenorhabditis elegans</del>						
and methods of using same						
JOURNAL Patent: WO 0161006-A 36 23-AUG-2001;						
Basf Corporation (US)						
FEATURES						
Location/Qualifiers						
Source 1..1394						
/organism="Caenorhabditis elegans"						
/db_xref="taxon:6239"						
BASE COUNT 347 a 317 c 270 g 460 t						
ORIGIN						

Query Match 100.0%; Score 1388; DB 6; Length 1394;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atggttaataatcaaccagatcgaaacacattatccctgttgagcaggaagcatttcccaagagac	60
Db	1	ATGCTAATAATCAACCGATCGAACACCTATGCCGTGTAGCAGGAAGCATTTCCCAAGAGAC	60
Qy	61	aagtacaatatgtctactggctcgctcattcttctgttgattcgaggattctctcgcattgg	120
Db	61	AAGTACAATAATGTCTACTGTGCTCGCTCATTTCTGTGGATTCGTCTGCCATGG	120
Qy	121	aatatgttcattactatcgccctcgagattattatgtgaattattgttcaaacccgattggc	180
Db	121	AATATGTTCATTACTATGCGCCCTGAGTATTATGTGAATTAATGTGTTCAACCCGGATGGC	180
Qy	181	gtggagacatggatttcgaaagaattcatgggatctctttgacgattggctcacaacttcca	240
Db	181	GTGGAGACATGGTATTGGAAGAATTCAATGGATCTTTGACGATTTGGCTCACAACTTCCA	240
Qy	241	aacgaagcattaatgttttcaacactgttcactcattattatgctgctccctgattacgcg	300
Db	241	AACGAAGCATTAATGTATTTCAACCTGTTCCTCATTTATTTGCTGCTCCCTGATACCGC	300
Qy	301	gtcttgcctcggttcttcaacatcgtaacctgacaatcattctcatctcgtcatt	360
Db	301	GTCTTGTCTCGGTTTGTCTCAACATCGTCAACCTGCAATCATTCATCTCATCTCGTCAAT	360
Qy	361	gtctggagccactgaagattccatgctcctgggttttcttggttaacctcttggaatggcg	420
Db	361	GTCTGGAGCCCACTGAAGATTCCATGCTCCTGGTTTCTGGGTAACTCTTTGGAATGGCG	420
Qy	421	acttcaatcaatttagcaatgggctatatgaaaactcgtgttatgagtggtggtggat	480
Db	421	ACTTCAATCAATTTTAGCAATGGGCTATATGAAACTCGGTTTATGGAGTTGGTGGCGAT	480
Qy	481	ttccgcacacctacattggcgctctcttgattggaacaacatttgcgattctgata	540
Db	481	TTTCCGCACACCTACATTGGCGCTCTCTTGATTGGAACAACATTTGCGGATTCCTGATA	540
Qy	541	acggttgtaaaatcgagtgacattttctggaatgatgagcctaaacttgttgcaatc	600
Db	541	ACGGTTGTGAAATCGGAGTGAACCTATTTCTGAATGATGAGCCTAAACTTTGTGCAATC	600
Qy	601	gtctatttcggcatatggttggtgatcctctgtgtggtgcaattgacattttctttatc	660
Db	601	GTCATTTCGGCATATCGTTGGTGATCCTCTGTGTGTGCAATTGACATTTCTTTATC	660
Qy	661	acaaagcaagatttctaccactatcacatcaaaaagaaatggaattcgcgaaaaggcg	720
Db	661	ACAAAGCAAGATTTCTACCACTATCACATCAAAAAGAAATGGAATTCGCGAAAAGCGG	720
Qy	721	gaaaccagacagcgtctccattccattcttttgaccacattcacaactgtttatgggcaa	780
Db	721	GAAACCGACAGCGCTCTCCATCCATTTCTTGGACCACATTCACAAACTGTTATGGCAA	780
Qy	781	ctcttcaattgttggtctgtctgttgcggttactctcacaatcttccctgttatgatgcc	840
Db	781	CTCTTCAATGTTGGTTCTGCTTGGCGGTACTCTCAAACTTTCCCTGTTATGATGACC	840
Qy	841	gttaccactgtagagattccggcttctcaacaaaattatgctgaaacgatgaaatc	900
Db	841	GTTACCACCTGAGAGATTCGGGCTTCTTAACAAAATTTATCTGAAAACGATGAATC	900
Qy	901	tacatttgcacaagtttctctgcttctcaatttctgctgctgagttggtatccatagtt	960
Db	901	TACACTTTGCTCAAAAGTTCTCTCTCTTCAAAATTTGTTGCTGCGATTGGATCCATAGTT	960
Qy	961	gttcccaagattcactggcgacaccccggttaccctcaaatgtccataactctgctgct	1020
Db	961	GCTTCCAAGATTCACTGGCGACACCCCGTTACCTCAAAATTTGCCATAAATTTGCGTGCT	1020

Qy	1021	cttttoattccattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	1080
Db	1021	CTTTTCATTCCATTCT	1080
Qy	1081	ttctttgagtcactgacacatttttctgattggttgaattgcatgtcttttttccatgga	1140
Db	1081	TTCTTTGAGTCTACTGACATTTTGTGATTTGTTGGAATTTGCATGCTCTTTTTCACATGA	1140
Qy	1141	tactcagcgtctcggcaatgggatatacactcccaaacgctcgtgccatctcactactcaaga	1200
Db	1141	TACTCTACGCGCTCTGGCAATGGGATACACTCCAACGCTGCGCATCTCTACTACTCAAGA	1200
Qy	1201	tttcggtcagcttccggtttgcactcttattggttggccttctccacggtggcctgtgg	1260
Db	1201	TTTCCGCGCTACGCTTCGGTTTGCACTTATGTTGGCTTCTCACGCGTGGCTGTGG	1260
Qy	1261	ccggttattgagcacttcgtggacaagccaagtattcttataaattatttatagcatta	1320
Db	1261	CCCGTTGTTATTGAGCACTTCGTGGACAAGCAAGTATCTTATAAATATTATAGCATTA	1320
Qy	1321	gagtatactgttataattgttt	1380
Db	1321	GAGTATACTGTTTATGTTGTTTATTAAGCTGTGGAATAAATAATATTATTAATAAAA	1380
Qy	1381	aaaaaaaa 1388	
Db	1381	AAAAAAAA 1388	

RESULT 2  
U55376/c 31049 bp DNA linear INV 09-AUG-2001  
LOCUS Caenorhabditis elegans cosmid F16H11, complete sequence.  
DEFINITION U55376  
ACCESSION U55376  
VERSION U55376.1 GI:1280130  
KEYWORDS HTG.  
SOURCE  
ORGANISM

Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE  
1 (bases 1 to 31049)  
The C. elegans Sequencing Consortium.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2016 (1998)  
MEDLINE 93069613  
REFERENCE 2 (bases 1 to 31049)  
Wu, X.  
TITLE The sequence of C. elegans cosmid F16H11  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 31049)  
Waterston, R.  
AUTHORS Unpublished  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 4 (bases 1 to 31049)  
Waterston, R.  
AUTHORS Unpublished  
TITLE Direct Submission  
JOURNAL Submitted (18-APR-1996) Robert Waterston  
REFERENCE 5 (bases 1 to 31049)  
Waterston, R.  
AUTHORS Direct Submission  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA  
Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rwenematode.wustl.edu and jes@sanger.ac.uk

X wst good data

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46C11, 200 bp overlap; 3' cosmid is K02G10, 2500 bp overlap. Actual start of this cosmid is at base position 1 of CELF16H11; actual end is at 10112 of CELK02G10.

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

#### FEATURES

##### source

1. .31049

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="X"

/clone="F16H11"

/complement(1239..2839)

/gene="F16H11.3"

/complement(join(1239..1368,1421..1617,1669..1878,

1932..2131,2180..2466,2516..2649,2693..2839))

/gene="F16H11.3"

/note="weak similarity to HNP36 protein"

/codon\_start=1

/product="Hypothetical protein F16H11.3"

/protein\_id="AAA98003.1"

/db\_xref="GI:1280131"

/translation="MTVEINRNTYAVQEAAPPRDKYNIVYWLVLVGVLLPWNMTI  
TIAPEYVNTYWEKPDGVTWYKEFGSLTASQPNASINVENFLIAGPLIYRVF  
APVCFNVLITILVILEPTEDSWFETWLTGMATINSNGLIYENSVGVFAD  
FHYTGALLIGNCGLLTVKIGVTYFLNDEPKLVIAVFIGISVILLVCAIALF  
FITKODPYHHQGMIEIREKAETDRPSILMTFTNCQGFNVWFCAVLTIFP  
VMVTYTRGSLNKYMSNDEIYTLTLTSLFNLFAAIGSIVASIKHWPTPRYLF  
AIIILRALFIPFFCNRVQTPRAYVPFESTDIFVIGGIAMSFSGHYSALAMGYTEN  
VPSHYSRFAQLSVCTLMVGLLTGLMAVVIHFVDKPSIL"  
complement(7137..7683)  
/gene="F16H11.4"

/note="ceh-1"

/complement(join(7137..7253,7307..7456,7505..7573,

7621..7683))

/gene="F16H11.4"

/note="C. elegans unknown homeobox gene ceh-1 (GB:X52810)"

/codon\_start=1

/product="Hypothetical protein F16H11.4"

/protein\_id="AAA98004.1"

/db\_xref="GI:1280132"

/translation="MRRARAFYEQLVLENKFKTSRYLSVVERLNLATLOLSEIQ  
VKIWFQNRRTKWKHNPDQANTPQPPSDEQIQIQLPANPITSFSSLLPPIISF  
ANGVQLQSTIPGLFNLNLQILMPQNVGFN"  
complement(10470..14006)  
/gene="F16H11.5"

/note="nhr-45"

/complement(join(10470..10481,10529..10624,10701..10861,  
10907..11025,11086..11396,11509..11613,12150..12431,  
13147..13430,13366..13631,13796..14006))  
/gene="F16H11.5"

/note="contains a zinc finger of the C4-type; coded for by  
the following C. elegans cDNAs: yk347g3.3, yk347g3.5"

/codon\_start=1

/product="Hypothetical protein F16H11.5"

/protein\_id="AAA98005.1"

/db\_xref="GI:1280133"  
/translation="MTTHFSINRNLTFSETTSKISMSCLVCETDAHQHGFIRCCRAC  
AAPFRRLTMNLRYKCFDRKCEVSNFKRYSRCRCRYEKCVRMGRKDNLTSPVSVVQ  
NKSETKDSSCNETDSIGHSPSSLESVTCPPSPFEOQOQONQDPPLLMGNNGTFVQ  
EVASDAYQPSNDMHQNFSPHRTVIEVGRAVMTPTFPQPPHHPHPPDFQYTDLLS  
TDEQNSAIPSSFTQYNQPOEYNESGNQMFAMAQAQATQDLHVNVQNFVPTIID  
AMFQPPDAPCOLQDILTLCCQALLAYREHNKQMPDQDKMIENVPLDMNFRNHYI  
EIEHIAFCMSIRVAPQLPKDKWIFKHEWTRFYEIDRCFATCQRQGYNLTDGRGLT  
LNGOIIINGFISVNLKLEISDMATQVRNFKGMDKPLRIINFPKQLQPTPEYELMYM  
MMSIMWSVNLPGITDRTDRISEKVELRLAEDLHTYYAEQYDNNNPNYAGRIIRLSII  
ASAYDEITERKREDSQVSKTFNFKNDFFFSDLTDFPCLN"  
16440..16613  
/gene="F16H11.2"

gene

16440..16613

CDS

/gene="F16H11.2"

/codon\_start=1

/product="Hypothetical protein F16H11.2"

/protein\_id="AAA98006.1"

/db\_xref="GI:1280134"

/translation="MTQDRTERERERERERERERERERERERERERERERE  
RERERESNRKR"  
21734..26707

gene

21734..26707

CDS

/gene="F16H11.1"

/join(21734..21787,21843..21909,21958..22079,24004..24109,  
24156..24295,24739..24809,24855..24966,25036..25127,  
25173..25257,25305..25397,25441..25536,25584..25751,  
26297..26394,26451..26593,26649..26707)

/gene="F16H11.1"

/note="similar to melibiose carrier protein

(thiomethylgalactoside permease II); coded for by the

following C. elegans cDNAs: cm01e2, cm21e6"

/codon\_start=1

/product="Hypothetical protein F16H11.1"

/protein\_id="AAA98007.1"

/db\_xref="GI:1280135"

/translation="MTDEETGDPVVRSCDDIVIPRRSIRPDHIDISTKKIARFIGH  
FYNDLCASMMFTYFMIYNEKVLKFOSSRAGMLIGQVTDALSTPLVGIFSDSNILPA  
CFDKIGRMSHLIGTIVLSLSPFMFNKCLCKSTSEWLKVLWVFPIINVFOFGWA  
SVQIHLALIPELSPASRATMSRLRYAVTIANLSYFALWLLSESTGHTSIGPW  
DFSHERAGLWLVVLTGTVAFVFAFTREPTNYRFRSLSSDASRLVRHMTSWF  
GHVQFIQALLYMLSRILINISQVFFPYITMNTQYKYYAILPWHVYLLSSFSVSW  
NSLPVLSKKILYSFGLASGMLSCVMDLPQWKIYALAVGIIAQAILLITLSLS  
ITADLNNKTESGAFVYGAWSFFDKLSNGIAYQLIETWTPAYDALKPHEVSAIFRRV  
MVFPVPGCLVLAFLLVLSLAFPIKERRRARRPEQAIMEQDDDIYPEIVE"  
BASE COUNT 10124 a 5100 c 5591 g 10234 t  
ORIGIN

Query Match 72.9%; Score 1012; DB 3; Length 31049;

Best Local Similarity 82.0%; Pred. No. 1.9e-259;

Matches 1371; Conservative 0; Mismatches 5; Indels 296; Gaps 6;

Qy 1 atggttaataacaaccgatcaaacacctatgccgttgagcagaagcattccaagac 60

Db 2839 ATGGTAATAACAACGATCGAACACCATGCGGTGAGCAGGAGCATTTCCAAGAC 2780

Qy 61 aagtacaattgtctactgctcgtcattctgttgattcggagttctcttcgcatgg 120

Db 2779 AAGTACATATGTCTACTGGCTCGTCATCTTGTGGATTCGGAGTCTTCTGCCATGG 2720

Qy 121 aatagtcttactatcattcgccctg----- 145

Db 2719 AATAGTTTCATTACTATCGCCCTCGAGGTTTGCAACTAGATAATTTTCACTAATAGAGTT 2660

Qy 146 -----agttattatggaattattgttcaaacccgatggcgtggagacattgtattc 197

Db 2659 CCGTGTTCAGTATTATGTGAATTATTGTTCAACCGGATGCGGTGGAGACATGTTATTC 2600

Qy 198 gaaagaattcggatcttgacgattggctcacaacttccaaacgaagcattaaagt 257

Db 2599 GAAAGAAATCATGGGATCTTTGACGATTGCCCTCACAACTTCCAAACGCAAGCATTAAATGT 2540

Qy 258 ttcaacctgttctcattattgc----- 281

Db 2539 TTTCAACCTGTTCCCTCATTAATTCGCTAAAGTTACAGTTCTTCTTAAACTCTCAAAAATCAAT 2480  
QY 282 -----tggtccctgatctaccggtcttggctcgggttgcttcaacatcg 328  
Db 2479 CAATATTTTTAGTGGTCCCTCGATCACCGGGTCTTGGCTCCGGTTGCTTCAACATCG 2420  
QY 329 tcaacctgacaatctctcaccctcgctcaatgttctggagccacctgaagattccatgt 388  
Db 2419 TCAACCTGACAATCATTTCTATCTTCGTCATGTGTTCTGGAGCCCACTGAAGATTCATGT 2360  
QY 389 cctgtgtttcttgggtaaccttgggaatgggagacttcaatcaatcttttagcaatgggctat 448  
Db 2359 CCTGGTTTTCTGGGTAACTCTTGAATGGCGACTTCAATCAATTTTAGCAATGGGCTAT 2300  
QY 449 atgaaaaactcgggttataggatgggtggcgattttccgcacacctacattggcgctctct 508  
Db 2299 ATGAAAACCTCGGTTTATGGAGTTTTTGGCGATTTTTCCGCACACCTTACATTTGGCGCTCTCT 2240  
QY 509 tgattggaaaacaacatttgcggatttctgataaacggttggtaaaatcggagtgacctatt 568  
Db 2239 TGATTTGGAACACAAATTTGCGGATTTGCTGATAACGGTTGTGAAAATCGGAGTGACCTATT 2180  
QY 569 -----ttctggaatgatg 580  
Db 2179 GTAAGTAATTTTCAATTAATTTGCTTTAAAAACTATAACATTTTGTAGTTCTGTAATGATG 2120  
QY 581 aqctcaaaactgttgcacatgctctatttccggcatactgttggatcacttcttgggtgtg 640  
Db 2119 AGCCTAAACTTGTTCGAATCGTCTAATTCGGGATATCGTTGGTGATCCTCTCGGTGTG 2060  
QY 641 caattgcactttcttcatcaaaagcaagatttctacacatcaccatcaaaaaggaa 700  
Db 2059 CAATTGCATTTTCTTTATCAGAAAGCAAGATTTCTACCACTATCACCATCAAAAAGGAA 2000  
QY 701 tggaaattcggaaaagcggaaaacccgacagacgtctccatccatttcttggaccacat 760  
Db 1999 TGGAAATTCGGAAAAGCGGAAACCGACAGACCGCTCTCCATCTCTTTGGACCACAT 1940  
QY 761 tcacaac----- 768  
Db 1939 TCACAAACGTGAGTTTTTAAATCAATTTAGCTTCAAAATAAATATGTGTGACTAATTTCA 1880  
QY 769 -tggtatgggcaactcttcaatgtttgttctgttccgttactctcaaatcttccc 827  
Db 1879 GTGTTATGGCAACCTTCAATGTTGTTGTTCTGCTTTGCGGTTACTCTCACAATCTTCCC 1820  
QY 828 tgttatgatgaccttacacactcgttggagattccgggtctcctaaacaaaattatgtctga 887  
Db 1819 TGTTATGATGACCGTTACCACTCGTGGAGATTCCGGCTTCCTAAACAAAATATATGCTGA 1760  
QY 888 aaacgatgaatctacactttgctcacaaagtttctcgtcttcttcaatttggctgcgat 947  
Db 1759 AAACGATGAATCTACATTTGCTCACAAAGTTTCTCGTCTTCAATTTGTTGCGCTGGCAT 1700  
QY 948 tggatccatagttgttccaaagattcactg----- 977  
Db 1699 TGGATCCATAGTTGCTTCCAAGATTCACTGGTAAATTTGAATTCAGAAATAACATTGCT 1640  
QY 978 -----gcagacaccocggttaccctcaaatattggcataatcttgcg 1016  
Db 1639 TGAATPATCATCAAAATTTGCGAGCCGACACCCCGTTACCTCAAAATTTGCCATAATCTTCCG 1580  
QY 1017 tgccttttcttccattctctctctcgaactatcgttccagacgcatgcttctcc 1076  
Db 1579 TGCCTCTTTTCATTCATTCCTCTCTCTGCTGCAACATATCGTTCAGAGCCGCTGCTATCC 1520  
QY 1077 tgtttcttcttgagtactgacatttttctgtgattgggtggaattggcattgtcttttccaa 1136  
Db 1519 TGTTTCTTTTGTAGTCTACTGACATTTTGTGATTTGGTGAATTTGCCATGCTCTTTTTCACA 1460  
QY 1137 tggatacctcagcgctcttggcaatgggatacactcaaa----- 1175  
|||||

Db 1459 TGGATACCTCAGCGCTCTGGCAATGGGATACACTCCAAGTGAGTTACAATTCGGTTGTT 1400  
QY 1176 -----cgctgaccatctcactactcaagatttg 1204  
Db 1399 ACTATGAATTTGGAATCTTTAAATTTTATAGCGTGTGGCATCTCTACTACTCAAGATTG 1340  
QY 1205 ccgctcagctttcccttggcactcttattggtggcctctctcaccggtggcctgtgcccg 1264  
Db 1339 CGGCTCAGCTTTCGCTTGGCACTCTTATGTTGGCTTCTCACCAGTGGCTGTGGCGG 1280  
QY 1265 ttgttattgagcaactcgtggacaagccaagtattcttataataattattagcattagagt 1324  
Db 1279 TTGTTATTGAGCACTTCGTGGACAAGCAAGTATCTTATAAATATTATAGCATTAGACT 1220  
QY 1325 atactgttatgtgttttttattaaactgtggataaataaataattattaaa 1376  
Db 1219 ATACTGTATATGTGTTTATTAACTGTGGAATAAAATAATATTATAA 1168  
  
RESULT 3  
AY058347 AY058347 2065 bp mRNA linear INV 16-OCT-2001  
LOCUS Drosophila melanogaster GH12067 full length cDNA.  
DEFINITION AY058347  
ACCESSION AY058347  
VERSION AY058347.1 GI:16182780  
KEYWORDS FLI CDNA.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 2065)  
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,  
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,  
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,  
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.  
TITLE Direct Submission  
JOURNAL Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
COMMENT Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, please visit our location  
and relationship to other sequences, including its location  
(http://fruitfly.berkeley.edu) or send email to  
cdna@fruitfly.berkeley.edu.  
Location/Qualifiers  
1. .2065  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="26E2-26E2"  
1. .2065  
/gene="BEST:LD04971"  
/note="alignment with genomic scaffold AB003614"  
/db\_xref="FLYBASE:FBgn0026585"  
347..1723  
/gene="BEST:LD04971"  
/note="Longest ORF"





Best Local Similarity 45.48; Pred. No. 3.6e-13; Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;									
Qy	47	catttccaagagacaag	tacaattgtctactggtctgctgattctctgttggattcggag	106					
Db	59747	CAGCGCCAAAGATAAT	TCCTTATCTGTTCCTTATATCTCTGCTGCACGCGGTGGCA	59688					
Qy	107	tctctgcgaatgaatat	gttctactactatcgcccc-----tgagtattattgtgaatt	160					
Db	59687	CGCTGATGCCATGGAACAT	GTGTTTCTATCAGCGCCAAAGTCTATTTTCGAGGACTCAAGTTTCG	59628					
Qy	161	attggttcaaacggatggcgt	gagacatggtattcgaaagaattcatggaattcttga	220					
Db	59627	GACCAAAACACACCGTGGCC	AGGAAGTGTACCGCACCCATTTTCATGCAAAATATGG	59568					
Qy	221	cgattggctcacaacttccaa	cgaagcattaatgttttcaacctgttccctcattattg	280					
Db	59567	GCTTTGGCTCGCAGATTC	TCAACCTGGTCTTCAACTGGCTGAACATCTTTGTCAACTTTG	59508					
Qy	281	ctgtgcccctgatctaacgc	gtcttctgcggtttgttctcaacatgctcaacctgacaa	340					
Db	59507	CGCGCGACCTGACCACCG	GAATCGTCTACAGCATCATCTTCGAAATGTCATTCTGCTGG	59448					
Qy	341	tcatttcatcctcgtcattg	tcttctggagcccaactgaagattccatgctcctgggttttct	400					
Db	59447	TTACCATTAATTTTGGCC	ATGCTCG-----ACTCCTCGCAGTGGCGGCGTCTTCTTCT	59394					
Qy	401	gggtaactcttggaaaggc	gacttcaatattagcaatgggctatatgaaactcgg	460					
Db	59393	GGACACCATCGTGTGCAT	TGTGTTAAACGTGTGCAACGGCATCTACCAACACCA	59334					
Qy	461	tttatggagttgttgpgatt	ttccgcacacctacattggcgctctcttgaattggaaca	520					
Db	59333	TATACGGAATAGTGGCAT	CGTTGCCATCAAGTACACAGGCGCGTCTGCTGGGCTCCA	59274					
Qy	521	acatttgcgaattcgtata	caggttgtgaaaaatggagtgacctattttctggaatgatg	580					
Db	59273	ACATCAGCGGTGTCTTACC	ACCGCCATGCGCTTGATATGCGGAGATCTTCCTCGTCCA	59214					
Qy	581	agcctaaactgttgcgaat	gctctatttcggcgcatactggttggatcctctgtgtgtg	640					
Db	59213	AG--AGGACTTCGGCCAT	CTACTACTCTGTCGACCGCATCTCTAGTGTGCTCTGCT	59157					
Qy	641	caattgcacttttctttat	caagaagatttctaccacataccacatcaaaaaggaa	700					
Db	59156	TCGACACGTACTTTGCG	CTGCGTGAACAAAGTTCTTCCGCGACTACGAGACCATTAGCC	59097					
Qy	701	tggaatttcgaaaggcgaa	accgacagacggtctccatctcttttggaccacat	760					
Db	59096	GGACGACGAAAGAAATCT	GATTTAAGGCCCGACGTGAACGTGCCTTTACTGGCGATCT	59037					
Qy	761	tcacaaactgttatggc	aaactcttcaatgtttggttctgtcttttgcgcttactctcaca	820					
Db	59036	TCAGAAAGACGCGCCCA	CTATTCACATCTTCCCTACGCTTTTCTGTCAGCTGTGCG	58977					
Qy	821	tcttccctgttatgatgac	cggttaccactcgtgagattccgggttctcctaaacaaatta	880					
Db	58976	TTTTCCCGCGGATTTCAG	TGCAACGTGACCGCTCGACCCCAATTTTCGTA-----	58925					
Qy	881	tgtctgaaacgatgaaat	ctacacttgtctcacaagtttctcgttcttcaatttcttcg	940					
Db	58926	-----GTAGGCCCGGAC	TACTTTCACATGGTCACTTGTCTCGGACCTTTAACGCTTTTCG	58872					
Qy	941	ctcgattggatccatagt	tgtcttccaaagtctcactgcccagacccccggttacctcacaat	1000					
Db	58871	CAATGCTGGCGAGTTTA	CCACATCGTGGTGCATGCGCGGGTCCGAGGTTCTCTCGG	58812					
Qy	1001	ttgcataaattcgtgctct	tttttccatctcatttcttcttcttgcgaacta-----tc	1054					
Db	58811	TCCAGTGGTCTGCGTTC	GTGCGTTCCTCCCTGTTGTCATGTCGAACCTACGTTCCGC	58752					
Qy	1055	gtgtccagacgcgtgctt	atcgtttttcttggatctactgacatttttggattggtg	1114					

FEATURES

```
source
1..174376
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="26F-27X"
/clone="BACR25K24 (DI294)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
BASE COUNT 48681 a 38138 c 38499 g 49058 t
ORIGIN
Query Match 6.5%; Score 90.8; DB 3; Length 174376;
Best Local Similarity 45.4%; Pred. No. 4e-13;
Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;
Qy 47 catttccaaagagacaagtacaaatattgtctactggtcgtcgtcattcttggtagttcggag 106
Db 39300 CAGCGCCAAAAGATAAATCCCTTATCGTGTCTTTATATCTCTGCTGCAGCGCGTGGCA 39359
Qy 107 tctctctgcatggaatattgttcaattactatcgcccc-----tgagtattatgtgaatt 160
Db 39360 CGCTGATGCCATGGAAACATGTTTCATCGGGCAAGTCTTATTTTCGAGGACTTCAAGTTGC 39419
Qy 161 atttggttcaaacggatggtggtgagacatggtattcgaagaattcatggatcttga 220
Db 39420 GACCAAAACACCCGTGCCAGGGAAGTAGTTACCGACCCATTTTCATGCAAAATATGG 39479
Qy 221 egattggttcacaacttcacaacgcaagcattaatgttttcaacctgttcctcattattg 280
Db 39480 GCTTTGCTCGCAGATTCTCACTACCTGGCTTCAACTGGCTGAACATCTTTGTCAACTTTG 39539
Qy 281 ctgtgtccctcatatcacgcgtcttctgtcgcgtgttcttcaacatcgtcaacctgacaa 340
Db 39540 CGCGGCACTTACCACCGAATCGTCTACAGCATCATCTCGAAATGTCATTTCTGCTGG 39599
Qy 341 tcaattctcatctcgtcattgtctctggagccacgtgaagattcccatgctcgtgtttct 400
Db 39600 TTACCATATATTTGGCCATGCTCG-----ACTCTCCGATGGCGCGGCTCTTCTTCT 39653
Qy 401 gggttaactcttggaaatggcgacttcaatcaatttttagcaatggggtatataaaactcgg 460
Db 39654 GGACCACCATGCTGTCATGTTGTTGTTAAAGCTGTGCAACGGCATCTACAGAACACCA 39713
Qy 461 ttattgagattggtggcgatcttcgcgcacacctacaattggcgctctcttgattggaaca 520
Db 39714 TATACGGAATAGTGGCATCTGTTGCCATCAAGTACACAGCGCGCTGCTCTGGGCTCCA 39773
Qy 521 acatttgcgattgctgataacggtgtgtgaaaatcgagatgacctattttctgaaatgatg 580
Db 39774 ACATCAGCGGCTGTTCAACCGCCATGGCTTGAATGCGGTGAGATCTTCTCGTCCA 39833
Qy 581 agctcaaaactgttgcaatcgtctatttcgagcatacgttggtagtctctcttgggtgtg 640
Db 39834 AG--AGGACTTGGCCCATCTACTTCTGTCGACCGCCATCCTAGTGTGCTGCTCTGCT 39890
Qy 641 caattgaccttttctttatcaagaagcagattttcacacatatcacacatacaaaaagaa 700
Db 39891 TCGACACGATCTATTGTCGCTGCGCTGGAACAAAGTTCTTCCGCGACTTACGAGACCATTAGCC 39950
Qy 701 tggaaattcgcgaaaagcgaaaacgcagacagcgtctccatctcttttggaccacat 760
Db 39951 GGAGCAGGAAAGAAATCTGATTTAAGGCCCGACGTGAAGTGGCTTACGTGGCAGATCT 40010
Qy 761 tcacaaactgttatgggcaactcttcaattgttggttcgtgttgcgttactctcaca 820
Db 40011 TCAAGAAAGACGCGCCCACTTAATCAACATCTTCTCCTCAGCTTTTTCGTACGTTGTCGG 40070
Qy 821 tcttccctgttatgatgaccttaccactcgtgagattccgcgttctcctaacaataa 880
Db 40071 TTTTCCCGGGGATTTCACTGTCGAACGTGACCGCTCCGACCCCAAAATTTTCGTA----- 40120
```

```
Qy 881 tgcctgaaaaagatgaaatctacacttttctcacaagtttctcgtcttctcaattgttcg 940
Db 40121 -----GTAGGCCCGGACTACTTTACACTGGTGCATCTGCTTCGGACCTTTTACGTTCTCG 40175
Qy 941 ctgcgattggatcccatagttgcttccaaagattcactggtccgcacacccggtttacctcaaat 1000
Db 40176 CAATCTGGGCAGTTTAAACCACATCGTGGTGCATTAACGCGGTCCGAGGTTTCTCTGGG 40235
Qy 1001 ttgccaataatttggcgtgctcttcttcaattccattcttcttcttcttcttcttcttcttctt 1054
Db 40236 TGCCAGTGTGCTGTGGCTGTTCATCCCTCTTGTGTCATGTCGAACATCGTTCGCCG 40295
Qy 1055 gtgtccgaacgcgtgcttctctctgtttttttgttggtctactgacatcttttctgattgg 1114
Db 40296 CGGACTCGTGGCTCATTTGGCGGTATTCATCGACAACGATTTGGGTGTACTGGGGCATCG 40355
Qy 1115 gaattgccaatgtcttttttcaatgatactcactcagcgtctgtgcaatgggatacactccaa 1174
Db 40356 GCATCCCAATGGCTTACAGCTCCGGGTATCTCAGCTCTCTGGGCATGATGTACGCCCGC 40415
Qy 1175 acgtgtgccaatctcactactcaaatgttgcgcgtcagcttcttccgtttctgacattcttattg 1234
Db 40416 AGACGCTGCACACCAAAATACCAGACCACTGCGGGAATGTACGACGTGCGCATGCTGATCA 40475
Qy 1235 ttggcctctcaccggtggtcgtgt 1258
Db 40476 CGGGCATCTTCTCCGAGAGTGTGT 40499
RESULT 7
AE003614 259464 bp DNA linear INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 14200001386055 section 7
DEFINITION of 63, complete sequence.
ACCESSION AE003614 AE002690
VERSION AE003614.2 GI:10728613
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 259464)
AUTHORS Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor
Miklos G.L., Abril J.F., Agbayani A., An H.J., Basu A.,
Andrews P., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y.,
Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkov D.,
Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C.,
Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de
Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.N.,
Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C.,
Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S.,
Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S.,
Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z.,
Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J.,
Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J.,
Wei M.H., Iqbal M., Kalish F., Karpen G.H., Ke Z.,
Kenshin J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C.,
Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A.,
Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C.,
McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C.,
Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L.,
Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,
Nusskern D.R., Pacieb J.M., Palazzolo M., Pittman G.S., Pan S.,
Pollard J., Puri V., Reese M.G., Reinert K., Remington K.,
```



Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.-Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhang, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)  
20196006  
2 (bases 1 to 259464)  
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7297129.  
Location/Qualifiers  
1. 259464  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
join(538..1127,1190..1366,1421..>1745)  
/gene="CG9550"  
/product="CT27016"  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
<538..>1745  
/gene="CG9550"  
/map="26D7-26D9"  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
join(538..1127,1190..1366,1421..1745)  
/gene="CG9550"  
/note="CG9550 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
/protein\_id="AAF52397.1"  
/db\_xref="GI:7297130"  
/translation="MVSLSRTSLTVACGVYTYVLFIFLLPMLPDPVTLQSLV  
SYSTHGLRNLTLDRGGQPIRSMLTVFSGALTLDLNLAHPGQYQHYAPLIAYE  
SRSLAHEHRLDEALAGYNCYDKNSAEMIQWSPRFRFYGAQKICRSYSQEA  
CWDPETMAAYCKLYPFINNAVYLNRLNLASLLERKDLNLMLLVRDPRGTYSRNN  
NKWCSDRDCEAOELCGDMVSDYQMVETLTAQYPORFSTIIRYEDLFTQPBESIKQVFD  
FYGLPLERNTRIQKHPRSSYFPOSSRMKELSSQSAFEWISEMWVDDIGAVDVCQ  
AMDLGWRSIQDNRHFSPEFQIMDKAY"  
2069..>3327  
/gene="CG9551"  
/product="CT27018"  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence=not\_experimental  
<2069..>3327  
/gene="CG9551"  
/map="26D7-26D9"  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence=not\_experimental  
2280..3327  
/gene="CG9551"  
/note="CG9551 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence=not\_experimental  
/protein\_id="AAF52398.1"  
/db\_xref="GI:7297131"  
/translation="MSTGICSSSSTEEOQHLELSADLEACSLLGPODPSNVDTGTRR  
AKORLSLNGKRRSRMRANLIGICGVSSLCILLIATQHRPLTTPFNQSGAGROG  
ASVDGPNLSAKSAFYDRFQQQQQSSQATVYNLTATIDVLSAQSRILAEME  
NFEPYRGGAERLDMTPTETGTPVRSVVVTSWRSGSTFLGLDILNSIPGNFYHPELLD

mRNA

gene

CDS

mRNA

gene

CDS

mRNA

gene

FGIKQIRDPDDQELAVQNLKLLNCDYADMIDYLNFGKTHYTLFEHNRRLWDVCREFP  
RFCWRPFLTRFCRLFTIQSMKTVRLRLAQAEKLEQOR"  
join(<27409..27443,28126..28308,28377..28578,  
28641..>28772)  
/gene="CG13765"  
/product="CT33249"  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence=not\_experimental  
<27409..>28772  
/gene="CG13765"  
/map="26D9-26E1"  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence=not\_experimental  
join(27409..27443,28126..28308,28377..28578,28641..28772)  
/gene="CG13765"  
/note="CG13765 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence=not\_experimental  
/protein\_id="AAF52399.1"  
/db\_xref="GI:7297132"  
/translation="MVMPQHVVISPIILSSVRIVLLVRDPRTQMOSRRHRWCGGNEDC  
EDPRLVCQDLRDDYKTAEVLLKLYPSRFRTVYEDLSLSPSEMTQDILFYGLPFPDA  
VEEFLDTHTKVNIIGVSVSTRDSRSAPFHMQLKPEIKIQDVCTEAMDLGWYRI  
ENFNFTTQOTFDPVMVPPPT"  
complement(join(30215..30949,31026..31270,31333..33087,  
35664..35710,38587..38869))  
/gene="cli"  
/note="Nucleotide sequence of the Celera sequence differs  
from the published sequence for this transcript."  
/product="CT27024"  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
complement(<30215..>38869)  
/gene="cli"  
/note="CG9554"  
/map="26D9-26E1"  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
complement(join(30768..30949,31026..31270,31333..33087,  
35664..35710,38587..38658))  
/gene="cli"  
/note="cli gene product; Nucleotide sequence of the Celera  
sequence differs from the published sequence for this  
transcript"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
/protein\_id="AAF52400.1"  
/db\_xref="GI:7297133"  
/translation="MVTLMFYNYAAPRCGLIDKMIPEKVKRPKTDHTDTHRNLCLNL  
SQQQQQQQQQQTHQQQQQQQQSHQSHSTVLASNGPSSAGAGMGVGGGGGGG  
GGVGGVGGCPLGPQSQPLQPTIGSLASLGHYSGNANPNVNSSCSLATASFF  
AQSAGSSFTYQAGGTGGVSGEDVGATVMSHTHDGTGSAVKSRSRSPGOV  
HASLDGSAAGSNLYGCSSAPLDGAVAVNSVAATAAAVYDGKHDYIYVNSMQQ  
YTPPPFYSGYGTYYAATAARQAQMEFGAAAAYLTPSYAASNNNSQLYSPYAG  
YNNFQQDYGITYNEQIGNYSPANYSPYAVSSPSASAGHGFHVAASNNLSSEPTD  
THSTPTVQHTHSPSPSPISPTSGIGPLGNVSAATAAAALNSGGSSVGTAGSGG  
VATSKTPTGKTGRGRHQQPSPTBSTADTGNSEAVKPERFVMDLDELTLIIFH  
TLLSGSYANRYTKDHSLLMTAFRMEEMVNMADTHFFNEIECDVOHIDVSDSDN  
GODLSAYNEATDGPHTPPCAPNLCILPTGVRGVDMRMKLAFRYRKIKDIYNSYRG  
NVGTLGLPGREAWLQIRSEIEVATDNWATLAKLSMISQRENCVNVVLVSTQLAPA  
LAKVLLGLGGINIENIYSAHKIGTCEYRIVTRFGKRSYTVYVIGDNGNEETAAGA  
MNEFPWRISAHSDIRALYALDIMGFL"  
complement(join(<58570..59115,59172..59817,60078..>60118))  
/gene="CG9595"  
/product="CT27132"  
/db\_xref="FLYBASE:FBan0009595"  
/db\_xref="FLYBASE:FBgn0031829"  
complement(<58570..>60118)

[illegible]

268882.  
The end of this sequence (31750..31853) overlaps with the start of  
sequence Z68299.  
For a graphical representation of this sequence and its analysis  
see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F44D12)  
name=F44D12  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.

## FEATURES

source	Location/Qualifiers
gene	1..31853 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="IV" /clone="F44D12"
gene	join(3277..3413,3774..3951,3994..4239,4318..4603, 4647..4894,4945..5180,5225..5395,5449..5759,5805..6053, 6100..6802,6846..7020,7211..7306,7357..7425) /genes="F44D12.1"
CDS	join(3277..3413,3774..3951,3994..4239,4318..4603, 4647..4894,4945..5180,5225..5395,5449..5759,5805..6053, 6100..6802,6846..7020,7211..7306,7357..7425) /genes="F44D12.1" /note="Similarity to Mouse Tight junction protein ZO-1 (SW:Z01_MOUSE), contains similarity to Pfam domain: PF00595 (PDZ domain (Also known as DHR or GLGF)). Score=35.0, E-value=5.5e-07, N=2 cDNA EST yk16d12.5 comes from this gene cDNA EST yk45f4.5 comes from this gene cDNA EST yk35f8.5 comes from this gene cDNA EST yk129c3.3 comes from this gene cDNA EST yk129c3.5 comes from this gene cDNA EST yk360b12.3 comes from this gene cDNA EST yk360b12.5 comes from this gene cDNA EST yk352d5.3 comes from this gene cDNA EST yk352d5.5 comes from this gene cDNA EST yk324a9.3 comes from this gene cDNA EST yk324a9.5 comes from this gene cDNA EST yk219a4.3 comes from this gene cDNA EST yk219a4.5 comes from this gene cDNA EST yk218f7.3 comes from this gene cDNA EST yk218f7.5 comes from this gene cDNA EST yk220g3.5 comes from this gene cDNA EST yk252d9.5 comes from this gene cDNA EST yk357f4.5 comes from this gene cDNA EST yk388f10.5 comes from this gene" /codon_start=1 /protein_id="CAA92607.1" /db_xref="GI:3877158" /db_xref="SPTREMBL:Q20398"
gene	/translation="MEKEIREVEAKWMEKVRKASERDDIGKKIDFEFKIDSLOIQV HRSETYKMEIEANKRLRQYALQRDYETQWERSIVLEENSRQNEERDLQKELEK MKEDSEVALKOKLENTRLKLVMEETAQANKEAARLEATEAQNGARITEERD EALRCKQNTDAIDVINTWTHSQINLPFNKPLGMLGGRTDDGSIYHGPIYVQRI AHGSPDFNLTKLIDHIMWINDISVTDMSRVNGLSNCHIIHLVIRRSNCNKISDV CLPLNTVGLSGNVFINSCEPGAASRGLAPGVVHVMTPTVYDAKHAEMLIK SREPLVIGLIQSTKRGDNGKDKHROTIFSRWFSRNGSDKERTVYVAKNIDRSNDQV LLRQSLRWPQASPASPLRYGSLRAPYSSIDHTKLMLDALKRFRNAKTAHSE ISSSTAAPTALMSPSSVNEKOLVFAGTNEGVPVYVPKSTIFSTPTSPVTRLRQNS EQKRTWSQISHRGGDGRFSNTTSDARPSMHTPTSTIMEGKPHRRSAVYSP SHPVYPTTRDSSMSVMSNSIRLPSTFSNQYNNSTSLTGGSGVPPRHCISKD GDSFSTSTGSAHSAEGRSKKHYLPRGNHSESLVLRNSSQPRLVEVPRSDVKLC GGNIGILAEKISGLLEDLLISIDGNCVNTTLECAMNTLSADNSELATGLLYODG GDRNLRLGADGDSFFLRNIDRSMENKDELKCGDVVFDKTLMLGKTGRRAWK VDKRGQRREGAIPSTTVYQAIRANRYANFPFKAYEWKLDITVKRPVLFGAVY EPTQLMLVDESDFRESFVARESLDSEYALKDKVLDSKQNDVDYLYHVSTAH IMDITAQGLHCVLQVQSAIDRLKRCMPILVKIRFQSVKOLKDVNEHCIGEKISK EAKOLIEKDLKIEKLDGSGVTLPVPSHNNVSEMTTHAVLQLKKIIEDEQKKIYVQVK VDEE" complement(join(7556..7673,7717..7832,8096..8296, 8340..8510,8562..8718,8944..9111,9541..9638)) /gene="F44D12.2"

CDNA EST yk302c7.5 comes from this gene; CDNA EST yk278d6.3 comes from this gene  
CDNA EST yk278d6.5 comes from this gene; CDNA EST yk264e2.3 comes from this gene  
CDNA EST yk264e2.5 comes from this gene; CDNA EST yk262a3.3 comes from this gene  
CDNA EST yk262a3.5 comes from this gene; CDNA EST yk233c12.3 comes from this gene  
CDNA EST yk233c12.5 comes from this gene; CDNA EST yk232g4.3 comes from this gene  
CDNA EST yk232g4.5 comes from this gene; CDNA EST yk232f9.3 comes from this gene  
CDNA EST yk232f9.5 comes from this gene; CDNA EST

Query Match 6.1%; Score 84.8; DB 3; Length 31853;  
Best Local Similarity 57.0%; Pred. No. 1.4e-11;  
Matches 155; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
QY 283 ggtccctgattaccgcgctttgtctccggtttgtcttcaacatcgtaacacctgacaatc 342  
Db 12947 GGTTCATGATGCTTCGAGTTGTGGTCTCTCTCATTTGTAAGTGCATTTTGATTGGGCTC 13006  
QY 343 attctcatctcgtcattgtcttgagccactgaagattccatgctcgtgtttttctgg 402  
Db 13007 ATTGTTATTTTGGCAATTTTGTAAACACCATCTCCAGACAGTGTACATGTTTATATT 13066  
QY 403 gtaactcttggaaggcgacttcaatcaatttagcattgagctatgaaactcgatt 462  
Db 13067 GTTACTCTAATCATCATTAATGCTATGAATTTGGCCAATGAATTTCAAAATTCCTGTA 13126  
QY 463 tatggagtgtggtggttttccgcacacctacattgctcgtctcttctgattggaaacaac 522  
Db 13127 TACGGAATCGTCGGCGATTTTCGGGATAACTACATTAATCTCTTGGTTATTTGGAAACAAT 13186  
QY 523 atttcgcatgctgataacacggtgtgaaaat 554  
Db 13187 TTGTGTGGAGTTTACTTCAGTATTGAGTAT 13218

RESULT 9  
CEC47A4  
LOCUS CEC47A4 40664 bp DNA linear INV 24-JAN-2002  
DEFINITION Caenorhabditis elegans cosmid C47A4, complete sequence.  
ACCESSION Z82263  
VERSION Z82263.1 GI:3869220  
KEYWORDS HTG.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans  
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (sites)  
AUTHORS none.  
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium  
JOURNAL Science 282 (5396), 2012-2018 (1998)  
MEDLINE 99069613  
REMARK The C. elegans Sequencing Consortium.  
REFERENCE 2 (bases 1 to 40664)  
AUTHORS Mortimore,B.J.  
TITLE Direct Submission  
JOURNAL Submitted (11-NOV-1996) Nematode Sequencing Project, Sanger  
Institute, Hinxton, Cambridge CB10 1SA, England and Department of  
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:  
jes@sanger.ac.uk or rwenematode.wustl.edu  
COMMENT On Nov 13, 1998 this sequence version replaced gi:2528885.  
Coding sequences below are predicted from computer analysis, using  
predictions from Genefinder (P. Green, U. Washington), and other  
available information.  
Current sequence finishing criteria for the C. elegans genome  
sequencing consortium are that all bases are either sequenced  
unambiguously on both strands, or on a single strand with both a  
dye primer and dye terminator reaction, from distinct subclones.  
Exceptions are indicated by an explicit note.

[981006 dl] : Cosmid flipped  
IMPORTANT: This sequence is not the entire insert of clone C47A4.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we arrange for a small overlap between  
neighbouring submissions.  
The true right end of clone C47A4 is at 40664 in this sequence. The  
right left end of clone F52D4 is at 101 in this sequence. The true  
right end of clone F56F12 is at 9819 in this sequence. The start of  
this sequence (1..104) overlaps with the end of sequence 280218.  
The end of this sequence (40564..40664) overlaps with the start of  
sequence 282273.  
For a graphical representation of this sequence and its analysis  
see:- [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C47A4)  
[name=C47A4](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=C47A4)  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.  
FEATURES  
Location/Qualifiers  
1..40664  
/organism="Caenorhabditis elegans"  
/db\_xref="taxon:6239"  
/chromosome="IV"  
/clone="C47A4"  
/complement(join(3491..3973,4192..4254))  
/gene="C47A4.1"  
/complement(join(3491..3973,4192..4254))  
/gene="C47A4.1"  
/codon\_start=1  
/protein\_id="CAB62792.1"  
/db\_xref="GI:6562350"  
/db\_xref="SPTREMBL:Q9U3L9"  
/translation="MTIKLLKLENFIVTYPSETLLEMDRNGKLNKNGKGVNDDEVIK  
QIIDLNDVTGGYKRESIYDTLAWHTIIFPLTFYIKWTALWRAIAQEKERDDDA  
KLYLRKIGVSQMEFQKTYDIDDLDFRECWLKATWRAERDAQAEKQKQDGA  
RYKRYKRYMKNAGTISEVD"  
join(15737..15769,15884..16078,16580..16814,17751..18037,  
18821..18923,18970..19055,19306..19502,20056..20176)  
/gene="C47A4.2"  
join(15737..15769,15884..16078,16580..16814,17751..18037,  
18821..18923,18970..19055,19306..19502,20056..20176)  
/gene="C47A4.2"  
/note="contains similarity to Pfam domain: PF01733  
(Nucleoside transporter), Score=142.5, E-value=2.4e-39,  
N=1"  
/codon\_start=1  
/protein\_id="CAB62793.1"  
/db\_xref="GI:6562351"  
/db\_xref="SPTREMBL:Q9U3L8"  
/translation="MVIRDESISPVDPPLPTQDQNTDLASIVDAGSSDEHNELI  
PEDKGRIVFWILLNGIGVLLPNMFTTIAPOYVDYVFTVNGTATVADSFOSAMGV  
VAQVNLIVAIINVLIRGPLYRVLAPLAFNSLIVILVILVIFQPSDQARNWFY  
IVSLIIVMANASNGLLYQNSFFGMAADFPKYNNAVVGICNIGCTFTSVLAIVTLAF  
STQAEVVALIYFGLSLILEFLVSWMFCMKMCLQICVFLVYFVSLVFPVTLAF  
FQSYTFPNDVYAGIAVFLNFNFAAVGNAATVTFPGPRLIVPCVIRLLFPFPM  
FSNYLPHSRMTGVLFTNWEIFFCTLLAFTSGYFSSGLGMMYTPRVCPPEYSKLGOV  
SALSLVGLITAGVSFTYAITAMVNF"  
complement(join(28627..28707,28755..28832,28884..29024,  
29070..29267,29338..29466,29527..29850))  
/gene="C47A4.3"  
complement(join(28627..28707,28755..28832,28884..29024,  
29070..29267,29338..29466,29527..29850))  
/gene="C47A4.3"  
/note="contains similarity to Pfam domain: PF00149  
(Ser/thr protein phosphatase), Score=476.0,  
E-value=1e-139, N=1  
CDNA EST yk121b5.3 comes from this gene  
CDNA EST yk121b5.5 comes from this gene  
CDNA EST yk262b6.3 comes from this gene  
CDNA EST yk262b6.5 comes from this gene  
/codon\_start=1  
/protein\_id="CAB62794.1"  
/db\_xref="GI:6562352"

```

/db_xref="SPTREMBL:Q903L7"
/translation="MQNVVDSIIIDVLSASTHEKPLCKVITEERVVLKLDLALGVFK
AOKPMVFNAPIKVGDIGHQFDDLLRFLHFGWPTTANVFLGDDVDRGRFSIETIV
LLLAYVKFPCNFELLGRNHECEFNKTYGEEOCKRYOSVRMYAFOFVNWMLPLT
GLIATKILOMHGGISPLMTKESTLDLTKRIERTCEKEGLADLWADPISGLSGPMN
NORGACGGRSDVSLNCSEFQDLVLCRAHVVDQYFEFFAGRKLVTIIFSAPHYCCQF
DNCAAFMSDEKLQCSFEILLRPTTGRLEIREKPLKDETN"
gene
33294..33521
/genes="C47A4.5"
CDS
33294..33521
/genes="C47A4.5"
/codon_start=1
/protein_id="CAB62795.1"
/db_xref="GI:6562353"
/translation="MADKSAFVPVDAIGNHKNKNTDLDDIDDELFGKGGKPKTAEAPKA
PPPPAPAPAAASTGYQYKSKSTYOKSYAK"
BASE COUNT 13436 a 6918 c 6544 g 13766 t
ORIGIN

```

```

Query Match 5.1%; Score 71.2; DB 3; Length 40664;
Best Local Similarity 53.6%; Pred. No. 6.1e-08;
Matches 148; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 390 ctggttttcttggaactcttggaatggcgacttcaatcaatttagcaatgggctata 449
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17758 CTGTTCTACATTGTCATTATTAATCATAGTAAATGCGCAATGAATGCATCCAAAGTCTCTA 17817
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 450 tgaataacatcggttatggagttggtggcgattttccgcacacactacattggcgctctctt 509
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17818 CCAAAATCATATTTTCGGAATGCGTGCAGATTTTCCGCAAAATACATCGAACGCTGTGGT 17877
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 510 gattgaatacaacattgctgattgctgataacaggttgtaaaatcgagtgacctattt 569
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17878 TATTGGAACAAATATTGTGGCAATTCACATAGTGTCTTGGCAATTCGCAACTTTGGC 17937
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 570 tctgaatgatgagcgttaaaactttgtgcaatcgtctatttctggcgatattcgttggtgacct 629
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17938 ATTCTCAACAAAGCTGAACACTGTGCACCTATTATTATTGGAAATACACTTCTTATTCT 17997
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 630 tctggtgtgtaattgcaattctttctttattacaaaa 665
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17998 GTTTGTTGTCTTGTGTCATGGTGTCTGCAAGAA 18033
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 10
CEZK809/c
LOCUS CEZK809 25560 bp DNA linear INV 11-DEC-2001
DEFINITION Caenorhabditis elegans cosmid Zk809, complete sequence.
ACCESSION 268303
VERSION 268303.1 GI:1130662
KEYWORDS HTG; Hydrophobic nucleolar protein like; Peroxisome assembly factor like; Protein-tyrosine phosphatase.
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (sites)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 25560)
AUTHORS Dobson, R.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jess@sanger.ac.uk or rw@nematoe.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using

```

predictions from Genefinder (P. Green, U. Washington), and other available information.

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is not the entire insert of clone ZK809. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone ZK809 is at 1 in this sequence. The true right end of clone ZK809 is at 15121 in sequence Z68302.

The true left end of clone ZK792 is at 25461 in this sequence. The start of this sequence (1..113) overlaps with the end of sequence Z73427.

The end of this sequence (25461..25560) overlaps with the start of sequence Z68302.

For a graphical representation of this sequence and its analysis see: - <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=ZK809>

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

Location/Qualifiers

1..25560

/organism="Caenorhabditis elegans"

/db\_xref="taxon:6239"

/chromosome="IV"

/clone="ZK809"

join(1372..1376,1431..2001,2054..2551,2599..2803,2847..2854)

/gene="ZK809.1"

join(1372..1376,1431..2001,2054..2551,2599..2803,2847..2854)

/gene="ZK809.1"

/notes="contains similarity to Pfam domain: PF00102 (Protein-tyrosine phosphatase), Score=35.2, E-value=2.3e-13, N=1"

cDNA EST YK36596.3 comes from this gene"

/codon\_start=1

/protein\_id="CAA92641.2"

/db\_xref="GI:5824906"

/translation="MPNKGVGKSPSTKSSNNKRRPESEYEKTAHVHGKTKVE EKTVDVHGKTKNVVAKKVPKASVIVDKKASTTETTVPTVTKKSMLSISATYDRTVE EKEKPKPKKEKKEKSGEDASIKFKELRNQSEGVRDRDTEASCPNCKMQSHSRN RHTDYKYDHNRIIVQCKDDYINGSKINVPHTPSIYLVLPKMSDVADEFWRVV FEQCOTVHTIARPEELTNPAIDKLCOESGAWLYANGFFVNTKRVKKNADMFV VELLPGCSNAVMCSVLYHTYKPLFGPDGGAQIARAHOIAKNGNSPTVASNG SGNAALLTLIAVVEDQLTRGKEPKSEIVRTIREQRPQSDSYIQVYSLVMAFNLLIK TKLPSGHEVVSKIKKFNKATRSPLGSMKMTPEEL"

join(3252..3612,3660..3835,4567..4606,5066..5406)

/gene="ZK809.7"

join(3252..3612,3660..3835,4567..4606,5066..5406)

/gene="ZK809.7"

/note="Similarity to Rat peroxisome assembly factor-1 (SW:PAFL\_RAT), contains similarity to Pfam domain: PF00097 (Zinc finger, C3HC4 type (RING finger)), Score=16.2, E-value=0.00035, N=1"

cDNA EST YK560a11.5 comes from this gene"

/codon\_start=1

/protein\_id="CAA92640.1"

/db\_xref="GI:3881798"

/db\_xref="SPTREMBL:Q23601"

/translation="MLHFRYHOKRNILPNNTLHCTLPHYFSVFLSDTSFYLIFF SFYSIYLYFKMACRCILYDOLDADHVDSEIRSDSDVDVNVQMLPWARIFEKLR PELKYLEAVLWTHRSRSGASPCQEQMDIAYGTTPKLITSHFWFVIVPYLIRMSK LSGRIDIRMSYKIEAIYEIGLSLLHFLYLRASINVTQNRHMMHAFHDXVLLTYF FTKARQIRVKKOKLRDTSIGSSDIECVDCDKPSPVIMVQKCGHVCYCTIATS RNMTCPCLSENAPMEBFLAKKLKILIE"















**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query Match	Length	DB ID	Description
1	1388	100.0	1394	22	AAS12162
2	1386.4	99.9	1388	17	AAT18168
3	90.8	6.5	2060	23	ABL15935
4	90.8	6.5	4060	23	ABL15934
5	43.8	3.2	2354	23	ABL22074
6	43	3.1	8237	22	AAS46801
7	41	3.0	7657	22	AAS45477
8	41	3.0	7657	24	ABL34032
9	40.8	2.9	15732	22	AAS45389
					Caenorhabditis ele
					CORK potassium cha
					Drosophila melanog
					Drosophila melanog
					Drosophila melanog
					Tumour suppressor
					Chemically pretrea
					Human immune syste
					Chemically pretrea

XX  
PA (BADI ) BASF CORP.

XX Pausch MH;  
XX  
XX  
DR WPI: 2001-536570/59.  
DR P-PSDB; AAU07630.  
XX  
XX  
PT New polypeptide, a mutant potassium ion channel protein for improving  
PT inward potassium flux under acidic conditions -  
XX  
XX Example 14; Fig 9; 131pp; English.  
XX  
CC The invention relates to a mutant potassium ion channel protein, having  
CC four membrane spanning domains and two pore forming domains, comprising a  
CC mutation at the second pore forming domain. The expression of the mutant  
CC protein in a cell confers improved inward potassium flux and the ability  
CC to grow in the presence of potassium. Mutant proteins and their  
CC corresponding polynucleotide sequences can therefore be used to improve  
CC inward potassium flux into cells under acidic conditions by modulating  
CC the membrane potential using therapeutic agents. The sequences may be  
CC used to develop agonists and antagonists of potassium channel proteins in  
CC order to control pests such as nematodes and insects. This sequence  
CC represents a Caenorhabditis elegans DNA encoding the transmembrane  
CC potassium ion channel protein, CORK.  
XX  
SQ Sequence 1394 BP; 347 A; 317 C; 270 G; 460 T; 0 other;  
  
Query Match 100.0%; Score 1388; DB 22; Length 1394;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 atggttaataataacccgacgcgaacacccatgcggtgagcaggagcatttccaaagagac 60  
Dbb 1 atggttaataataacccgacgcgaacacccatgcggtgagcaggagcatttccaaagagac 60  
  
Qy 61 aagtcacaaatattgtactggtcgtcgtcatttctgttgattcggaggttcttctgcatgg 120  
Dbb 61 aagtcacaaatattgtactggtcgtcgtcatttctgttgattcggaggttcttctgcatgg 120  
  
Qy 121 aatatgttcactactatgcgcctcgtgatttatgtgaattatgttgcacacggatgac 180  
Dbb 121 aatatgttcactactatgcgcctcgtgatttatgtgaattatgttgcacacggatgac 180  
  
Qy 181 gtggagacatggtattcgaagaattcattggtatcttgaagattggtcacaactcca 240  
Dbb 181 gtggagacatggtattcgaagaattcattggtatcttgaagattggtcacaactcca 240  
  
Qy 241 aacgcaagcattaatgtttcaaacctgttcctcattatgtgtggtccctgatctaccgc 300  
Dbb 241 aacgcaagcattaatgtttcaaacctgttcctcattatgtgtggtccctgatctaccgc 300  
  
Qy 301 gtcttgctcgcggttctcacaacatcgtcaccctgacaacatcattctcatctcgtcatt 360  
Dbb 301 gtcttgctcgcggttctcacaacatcgtcaccctgacaacatcattctcatctcgtcatt 360  
  
Qy 361 gtcttgaggccactgaagattccatgctcgtgttttctggttaactcttggaatggcg 420  
Dbb 361 gtcttgaggccactgaagattccatgctcgtgttttctggttaactcttggaatggcg 420  
  
Qy 421 acttcaaatcaatttagcaaatgggctatatgaaacctcgggttatggaggttgggtggcgat 480  
Dbb 421 acttcaaatcaatttagcaaatgggctatatgaaacctcgggttatggaggttgggtggcgat 480  
  
Qy 481 ttctcgcacacctacattggcgctcctctgattggaacacacatttgcgatttgcgtgata 540  
Dbb 481 ttctcgcacacctacattggcgctcctctgattggaacacacatttgcgatttgcgtgata 540  
  
Qy 541 acggttgtaaaaatcggagtgacattttctgaatgatgatgacccataaactgttgcaatc 600  
Dbb 541 acggttgtaaaaatcggagtgacattttctgaatgatgatgacccataaactgttgcaatc 600  
  
Qy 601 gtctatttcggcatatcgttgggtgatcctctgtgtgtgcaattgcacttttctttatc 660  
Dbb 601 gtctatttcggcatatcgttgggtgatcctctgtgtgtgtgcaattgcacttttctttatc 660

Db 601 gtctatttcggcatatcgttgggtgatcctctgtgtgtgcaattgcacttttctttatc 660  
Qy 661 acaaaagcaagatttttacacactatcacacatcaaaaaaggaatggaatttcgcgaaaaagcg 720  
Dbb 661 acaaaagcaagatttttacacactatcacacatcaaaaaaggaatggaatttcgcgaaaaagcg 720  
Qy 721 gaaacccgacagaccgtctccatccatttttggaccacattcacaaaactgttatgggcaa 780  
Dbb 721 gaaacccgacagaccgtctccatccatttttggaccacattcacaaaactgttatgggcaa 780  
Qy 781 ctctccaatgtttgtgtctgtttgcttactctcactctcacaatcttccctgttatgatgacc 840  
Dbb 781 ctctccaatgtttgtgtctgtttgcttactctcacaatcttccctgttatgatgacc 840  
Qy 841 gttacacactcgtggagattccggtctctcctaaacaaattatctcgaacacgatgaatcc 900  
Dbb 841 gttacacactcgtggagattccggtctctcctaaacaaattatctcgaacacgatgaatcc 900  
Qy 901 tacatttgcacaaagtttctcgtcttctcatttgcctgtcgtcgtgattggtccatagtt 960  
Dbb 901 tacatttgcacaaagtttctcgtcttctcatttgcctgtcgtcgtgattggtccatagtt 960  
Qy 961 gcttccaagattcactggtcgcgacaccccggttacctcacaatttgcacataatcttgcgtgt 1020  
Dbb 961 gcttccaagattcactggtcgcgacaccccggttacctcacaatttgcacataatcttgcgtgt 1020  
Qy 1021 cttttcattcattcttcttctcgaactatcgttccagacgcgttcttactctctgtt 1080  
Dbb 1021 cttttcattcattcttcttctcgaactatcgttccagacgcgttcttactctctgtt 1080  
Qy 1081 tcttttgagtctactgacatttttggattggtggaattgccattgctttttcacatgga 1140  
Dbb 1081 tcttttgagtctactgacatttttggattggtggaattgccattgctttttcacatgga 1140  
Qy 1141 tactccaagcgtctggcaatgggatacactccaacacgtcgtgccatctcactactcaaga 1200  
Dbb 1141 tactccaagcgtctggcaatgggatacactccaacacgtcgtgccatctcactactcaaga 1200  
Qy 1201 ttgcccgtcagcttcggttgcactcttattggttggccttctcacccggtggccttgg 1260  
Dbb 1201 ttgcccgtcagcttcggttgcactcttattggttggccttctcacccggtggccttgg 1260  
Qy 1261 ccggtgttattgagcacttcgtggacaagccaagtattcttataaattattatagcatta 1320  
Dbb 1261 ccggtgttattgagcacttcgtggacaagccaagtattcttataaattattatagcatta 1320  
Qy 1321 gagtatactgttatattgttttttataagctgtggaataaaaaataattattataaaaaa 1380  
Dbb 1321 gagtatactgttatattgttttttataagctgtggaataaaaaataattattataaaaaa 1380  
Qy 1381 aaaaaaaa 1388  
Dbb 1381 aaaaaaaa 1388  
  
RESULT 2  
AAT18168  
ID AAT18168 standard; cDNA; 1388 BP.  
XX  
AC AAT18168;  
XX  
DT 16-OCT-1996 (first entry)  
XX  
DE CORK potassium channel gene.  
XX  
KW CORK; potassium channel; nematode; polyadenylation site;  
KW potassium dependence; Saccharomyces cerevisiae; potassium-agonist;  
KW potassium-antagonist; drug screening; nematocide; anthelmintic;  
XX cardiac disorder; ss.  
OS Caenorhabditis elegans.  
XX  
FH Key Location/Qualifiers

FT CDS 1..1305  
FT /\*tag= a  
FT /product= CORK potassium channel  
FT polyA\_signal 1359..1364  
FT /\*tag= b

XX WO9613520-A1.  
XX 09-MAY-1996.

PD 25-OCT-1995; 95WO-US14364.  
XX 31-OCT-1994; 94US-0323212.  
XX (AMCY ) AMERICAN CYANAMID CO.  
XX Pausch MH, Price LA;  
XX WPI; 1996-239450/24.  
XX P-PSDB; AAR92315.

Handwritten signature and date: 10/19/96

Potassium channel genes from *Drosophila melanogaster* and *Caenorhabditis elegans* - useful in assaying substances to determine effects on cell growth, and in inhibiting nematode and insect pests

Claim 13; Fig 9; 79pp; English.

The sequence encodes potassium channel CORK protein from *Caenorhabditis elegans*, with 2 pore-forming domains between transmembrane helix domains. The sequence has been isolated by complementation of the potassium-dependent phenotype of *Saccharomyces cerevisiae* CY162 (trk1-delta) on low-potassium medium. A consensus polyadenylation site is found in the 3'-untranslated sequence, and is followed by a tract of 15 consecutive A residues. The DNA may be inserted in a vector and expressed in a host cell to assay substances to determine effects on cell growth. Potassium-agonists or potassium-antagonists identified by this method may be used as nematocides, anthelmintics or in therapy of cardiac disorders, etc.

SQ Sequence 1388 BP; 341 A; 318 C; 269 G; 460 T; 0 other;

Query Match 99.9% Score 1386.4; DB 17; Length 1388;  
Best Local Similarity 99.9% Pred. No. 0;  
Matches 1387; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggtataatacaaccgatacaacacatccatgccgttgagcagggaagcatttcccaagagac 60  
Db 1 atggtataatacaaccgatacaacacatccatgccgttgagcagggaagcatttcccaagagac 60  
Qy 61 aagtacaattgtctactggtcgtcattctgtgtgattcgaggtcttctgcatgg 120  
Db 61 aagtacaattgtctactggtcgtcattctgtgtgattcgaggtcttctgcatgg 120  
Qy 121 aatgtttcattactatgccctcagtgatttatgtgaattatgttcaaacccggatggc 180  
Db 121 aatgtttcattactatgccctcagtgatttatgtgaattatgttcaaacccggatggc 180  
Qy 181 gtggagacatgtattcgaagaattcattggattttgacgattgggtcacaacttcca 240  
Db 181 gtggagacatgtattcgaagaattcattggattttgacgattgggtcacaacttcca 240  
Qy 241 aacgcaagcattaatgtttcaacctgtctcattatgtctgtgcccctgatctaccgc 300  
Db 241 aacgcaagcattaatgtttcaacctgtctcattatgtctgtgcccctgatctaccgc 300  
Qy 301 gtctttgctcggtttgcatacatcgtcaacctgacaacatcattctcatctcgtcatt 360  
Db 301 gtctttgctcggtttgcatacatcgtcaacctgacaacatcattctcatctcgtcatt 360  
Qy 361 gttctggagaccactgaagattccatgctcctggttttctggttaactcttgggaatggcg 420

Db 361 gttctggagaccactgaagattccatgctcctggttttctggtgaactcttgggaatggcg 420  
Qy 421 acttcaatcaatttttagcaatgggctatatgaaactcggtttctttaggttggtgagcat 480  
Db 421 acttcaatcaatttttagcaatgggctatatgaaactcggtttctttaggttggtgagcat 480  
Qy 481 ttccgcacacctacattggcgctctctgtgattggaaacacacatttgcggattgctgata 540  
Db 481 ttccgcacacctacattggcgctctctgtgattggaaacacacatttgcggattgctgata 540  
Qy 541 acggttgaaaaatcggagtgacctattttctgaatgatgagcctaaactttgtcaatc 600  
Db 541 acggttgaaaaatcggagtgacctattttctgaatgatgagcctaaactttgtcaatc 600  
Qy 601 gtctatttcggcattcgttggtgacacctctgtgtggtgcaattgcaacttttcttattc 660  
Db 601 gtctatttcggcattcgttggtgacacctctgtgtggtgcaattgcaacttttcttattc 660  
Qy 661 aaaaagcaagattttctacacctatcaccatcaaaaaagattgggaatttcgcgaaaaaggcg 720  
Db 661 aaaaagcaagattttctacacctatcaccatcaaaaaagatttcgcgaaaaaggcg 720  
Qy 721 gaaacggacagacgctcccatccatttttggaccacatttcaaaaactgttatggcaaa 780  
Db 721 gaaacggacagacgctcccatccatttttggaccacatttcaaaaactgttatggcaaa 780  
Qy 781 ctcttcaatgtttggtctgttgccttactctcaaatcttccctgttatgatgacc 840  
Db 781 ctcttcaatgtttggtctgttgccttactctcaaatcttccctgttatgatgacc 840  
Qy 841 gttaccactcgtggagattccggcttctcaaaaaatttatgtctgaaacgatgaaatc 900  
Db 841 gttaccactcgtggagattccggcttctcaaaaaatttatgtctgaaacgatgaaatc 900  
Qy 901 tacactttgctcaaaagtcttctcgttcttcaatttggctgctgagtgatggatccatgtt 960  
Db 901 tacactttgctcaaaagtcttctcgttcttcaatttggctgctgagtgatggatccatgtt 960  
Qy 961 gcttccaagattcactggtcggacaccccggttacctcaaatgttgcataatcttgcgtgct 1020  
Db 961 gcttccaagattcactggtcggacaccccggttacctcaaatgttgcataatcttgcgtgct 1020  
Qy 1021 ctttcttccattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1080  
Db 1021 ctttcttccattcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1080  
Qy 1081 ttctttgagtcactgacatttttggattggtggaattgcaatcttcttcttcttcttcttctt 1140  
Db 1081 ttctttgagtcactgacatttttggattggtggaattgcaatcttcttcttcttcttcttctt 1140  
Qy 1141 tacctcagcgtctcgtgcaatgggatacactccaaacgctggtccatctcaactcaaga 1200  
Db 1141 tacctcagcgtctcgtgcaatgggatacactccaaacgctggtccatctcaactcaaga 1200  
Qy 1201 ttgcccgtcagcttccggttggcactcttattggttggccttctcaccggtggcctgtgg 1260  
Db 1201 ttgcccgtcagcttccggttggcactcttattggttggccttctcaccggtggcctgtgg 1260  
Qy 1261 cccgtgttattgacacacttgggacgaagcacaatcttataataatttatacattta 1320  
Db 1261 cccgtgttattgacacacttgggacgaagcacaatcttataataatttatacattta 1320  
Qy 1321 gagtatactgttatgtgttttttataagcttggaataaaaaataattataaaaaa 1380  
Db 1321 gagtatactgttatgtgttttttataagcttggaataaaaaataattataaaaaa 1380  
Qy 1381 aaaaaaa 1388  
Db 1381 aaaaaaa 1388

RESULT 3  
ABLI5935























GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 22:11:17 ; Search time 89.48 seconds  
(without alignments)  
3810.229 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggttaataatcaacgctc.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 15 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCrUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	4.9	7218	1	US-08-232-463-14
2	38.6	2.8	1582	3	US-08-545-196B-10
3	38.6	2.8	1582	3	US-08-545-196B-12
4	36	2.6	1164	2	US-08-794-796-1
5	36	2.6	2621	2	US-08-553-619B-8
6	36	2.6	5852	1	US-07-867-106-2
7	35.8	2.6	1400	1	US-08-464-164-1
8	35.8	2.6	1400	1	US-08-338-057-1
9	35.8	2.6	1400	2	US-08-668-416-1
10	35.4	2.6	581	3	US-08-557-309B-22
11	35.4	2.6	581	3	US-08-834-306-22
12	35.4	2.6	581	3	US-08-993-674A-22
13	35.4	2.6	1303	2	US-08-793-410-11
14	35.4	2.6	3489	2	US-08-728-323A-1
15	35.4	2.6	32207	2	US-08-770-379-20
16	35.4	2.6	32207	4	US-08-757-669A-20
17	35.4	2.6	32207	4	US-09-230-371A-20
18	35.2	2.5	972	3	US-09-286-690-1
19	35	2.5	51259	3	US-08-781-891-209
20	34.8	2.5	289	4	US-09-007-005-17
21	34.8	2.5	289	4	US-09-244-796-17
22	34.6	2.5	991	4	US-09-377-648-7
23	34.6	2.5	1262	4	US-09-227-357-27
24	34.6	2.5	2238	1	US-08-742-011-1
25	34.6	2.5	3645	2	US-08-663-112-1
26	34.2	2.5	852	4	US-09-461-697-1
27	34.2	2.5	1991	2	US-08-415-593-40

28	34.2	2.5	2128	2	US-08-415-593-39	Sequence 39, Appl
29	33.8	2.4	519	1	US-08-339-582-1	Sequence 1, Appl
30	33.6	2.4	732	4	US-08-916-576B-5	Sequence 5, Appl
31	33.4	2.4	756	2	US-08-530-165-1	Sequence 1, Appl
32	33.2	2.4	174	2	US-08-378-235B-8	Sequence 8, Appl
33	33.2	2.4	248	4	US-09-007-005-32	Sequence 32, Appl
34	33.2	2.4	248	4	US-09-244-796-32	Sequence 32, Appl
35	33.2	2.4	277	4	US-09-007-005-3	Sequence 3, Appl
36	33.2	2.4	277	4	US-09-244-796-3	Sequence 3, Appl
37	33.2	2.4	1364	1	US-08-265-087-3	Sequence 3, Appl
38	33.2	2.4	1364	1	US-08-621-493-3	Sequence 3, Appl
39	33.2	2.4	1364	2	US-08-965-688-3	Sequence 3, Appl
40	33.2	2.4	1364	4	US-09-260-173-3	Sequence 3, Appl
41	33.2	2.4	2265	3	US-09-369-618-3	Sequence 3, Appl
42	33.2	2.4	2265	3	US-09-369-617-3	Sequence 3, Appl
43	33.2	2.4	6124	4	US-08-213-419B-3	Sequence 3, Appl
44	33	2.4	306	4	US-09-122-400B-9	Sequence 9, Appl
45	33	2.4	1939	1	US-07-715-751B-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232.463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935.313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTgpt-F1s  
US-08-232-463-14

Query Match 4.9%; Score 68; DB 1; Length 7218;  
Best Local Similarity 4.2%; Pred. NO. 1.5e-09;  
Matches 17; Conservative 234; Mismatches 149; Indels 0; Gaps 0;

Qy 907 ttgctcacagttctctgcttcaattgttcgtgctgagtgatccatagttcttc 966  
Db 1060 TTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1119  
Qy 967 aagattcactgacaccccggttacctcaaatggccataatctgctgctctttc 1026  
Db 1120 YY 1179  
Qy 1027 attcattctcttcttctgcaactatgctgcagcgtgtctatctcttcttt 1086  
Db 1180 YY 1239  
Qy 1087 gagtctactgacattttgtgattggtggaattgcatgctcttttccatagatactc 1146  
Db 1240 YY 1299  
Qy 1147 agcgtctggaattggatcacactccaaacgctggtccatctcactactcaagattgccc 1206  
Db 1300 YY 1359  
Qy 1207 gctcagcttctgcttgcactcttatgttgccctctcaccggtggtgctggtgcccgtt 1266  
Db 1360 YY 1419  
Qy 1267 gttattgagcacttgctgagacaagcagatctctataaa 1306  
Db 1420 YYYYYYYYYYYYYGTACCAAAATCTCTATCTCTTTAA 1459

RESULT 2  
US-08-545-196B-10  
; Sequence 10, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUNNICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FARACI, C. J.  
; REGISTRATION NUMBER: 32,350  
; REFERENCE/DOCKET NUMBER: 2121-110P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1582 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna

US-08-545-196B-10  
Query Match 2.8%; Score 38.6; DB 3; Length 1582;  
Best Local Similarity 63.4%; Pred. NO. 0.18;  
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1296 tatcttataaatatttatagcattagcagatgattgttattgttttttttaagctg 1355  
Db 1441 TATCTTCATATGTTTAAAGTATATATAATAAAATTTTAAATTTTAAAAA 1500  
Qy 1356 tgggaataaataattattataaaaaa 1388  
Db 1501 AAAAAAAAAAAAAAAAAAAAAAAAAA 1533

RESULT 3  
US-08-545-196B-12  
; Sequence 12, Application US/08545196B  
; Patent No. 6080577  
; GENERAL INFORMATION:  
; APPLICANT: MELKI, JUDITH  
; APPLICANT: MUNNICH, ARNOLD  
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/545.196B  
; FILING DATE: 19-OCT-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FARACI, C. J.  
; REGISTRATION NUMBER: 32,350  
; REFERENCE/DOCKET NUMBER: 2121-110P  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1582 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-08-545-196B-12

Query Match 2.8%; Score 38.6; DB 3; Length 1582;  
Best Local Similarity 63.4%; Pred. NO. 0.18;  
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1296 tatcttataaatatttatagcattagcagatgattgttattgttttttttaagctg 1355  
Db 1441 TATCTTCATATGTTTAAAGTATATATAATAAAATTTTAAATTTTAAAAA 1500  
Qy 1356 tgggaataaataattattataaaaaa 1388  
Db 1501 AAAAAAAAAAAAAAAAAAAAAAAAAA 1533



```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match 2.6%; Score 36; DB 1; Length 5852;
Best Local Similarity 55.6%; Pred. No. 1.9;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1265 ttgtttatgacactctgctggcaagccatcttataaatttatagcattagctg 1324
|| || | | | | | | | | | | | | | | | | | |
Db 2096 TTATTAAAAAATCAAAAAAACCACCAAGTAATATTATTTATAGGGGTTTTTT 2037

QY 1325 atactgttatgtgtttttttaaagctggtggaataaataattataaaaaaaa 1384
|| || | | | | | | | | | | | | | | | | | |
Db 2036 TTTTTTTTTTTTTTTTTTTTTTTCAAAAGTAAAAAAGAAAAAAGAAATA 1977

QY 1385 aaaa 1388
|||
Db 1976 GAA 1973

RESULT 7
US-08-464-164-1
; Sequence 1, Application US/08464164
; Patent No. 5614195
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; APPLICANT: Bumstead, Janene M.
; APPLICANT: Vermeulen, Arno N.
; TITLE OF INVENTION: Coccidiosis poultry vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5614195el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464.164
; FILING DATE: June 2, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match 2.6%; Score 35.8; DB 1; Length 1400;
Best Local Similarity 69.0%; Pred. No. 1.1;
Matches 49; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1318 tttagagtatactgttatgtgtttttttaaagctggtggaataaataattataaa 1377
|| || | | | | | | | | | | | | | | | | | |
Db 1326 TCAGCAAAATGCTTCTTAAATTATGTGTAATACTGCACGAGAATAAATAAATAA 1385

QY 1378 aaaaaaaaaa 1388
|||||
Db 1386 AAAAAAAAAA 1396

RESULT 8
US-08-338-057-1
; Sequence 1, Application US/08338057
; Patent No. 5795741
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; APPLICANT: Bumstead, Janene M.
; APPLICANT: Vermeulen, Arno N.
; TITLE OF INVENTION: Coccidiosis poultry vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; STREET: 1330 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/338,057
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 93.309078.9
; FILING DATE: 12-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

.....

Query Match 2.6%; Score 35.4; DB 2; Length 581;  
Best Local Similarity 61.3%; Pred. No. 0.89;

Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1296 tatctataaaattatagcattagctatctgtttatctgttttttttttaagctg 1355

Db 486 TTTTATATGCTCTTTTATTTGAGAAGATGCTGTTGTTGTTGTTTTCATCA 545

Qy 1356 tggataaaataattataaaaaa 1388

Db 546 AAAAAAAAAAAAAAAAAAAAAAAAAA 578

## RESULT 11

US-08-834-306-22

; Sequence 22, Application US/08834306

; Patent No. 6054135

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/834.306

; FILING DATE: 15-APR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 581 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-834-306-22

## Query Match

Best Local Similarity 2.6%; Score 35.4; DB 3; Length 581;

; Sequence 22, Application US/08834306

; Patent No. 6054135

Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1296 tatctataaaattatagcattagctatctgtttatctgttttttttttaagctg 1355

Db 486 TTTTATATGCTCTTTTATTTGAGAAGATGCTGTTGTTGTTTTCATCA 545

Qy 1356 tggataaaataattataaaaaa 1388

Db 546 AAAAAAAAAAAAAAAAAAAAAAAAAA 578

## RESULT 12

US-08-993-674A-22

; Sequence 22, Application US/08993674A

; Patent No. 6228372

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Smith, John M.

; APPLICANT: McNeill, Patricia D.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF

; NUMBER OF SEQUENCES: 81

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/993.674A

; FILING DATE: 18-DEC-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 581 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-993-674A-22

## Query Match

Best Local Similarity 2.6%; Score 35.4; DB 4; Length 581;

; Sequence 11, Application US/08793410

Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1296 tatctataaaattatagcattagctatctgtttatctgttttttttttaagctg 1355

Db 486 TTTTATATGCTCTTTTATTTGAGAAGATGCTGTTGTTGTTTTCATCA 545

Qy 1356 tggataaaataattataaaaaa 1388

Db 546 AAAAAAAAAAAAAAAAAAAAAAAAAA 578

## RESULT 13

US-08-793-410-11/c

; Sequence 11, Application US/08793410

; Patent No. 5955650

; GENERAL INFORMATION:

; APPLICANT: HITZ, WILLIAM DEAN

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA

; TITLE OF INVENTION: AND SOYBEAN PALMITOYL-ACP THIO-

; TITLE OF INVENTION: ESTERASE GENES AND THEIR USE IN

; TITLE OF INVENTION: THE REGULATION OF FATTY ACID

; TITLE OF INVENTION: CONTENT OF THE OILS OF SOYBEAN

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

; STREET: 1007 MARKET STREET

; CITY: WILMINGTON

; STATE: DELAWARE

; COUNTRY: USA

; ZIP: 19898

; COMPUTER READABLE FORM:

```

/ CLASSIFICATION: 4353
/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John P.
/ REGISTRATION NUMBER: 28,678
/ REFERENCE/DOCKET NUMBER: 52342
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 278-0400
/ TELEFAX: (212) 391-0525
/ INFORMATION FOR SEQ ID NO: 20:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 32207 base pairs
/ TYPE: nucleic acid

```

